Advanced Density Functional Theory Methods for Materials Science

Thesis by

Steven Demers

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2014

(Defended Nov 25, 2013)

Dedicated to GG, my rope across the abyss

and my parents who always gave everything.

Acknowledgments

For me, this thesis has been more of a journey than a destination, and as with any voyage it is the people you travel with that makes it possible and worthwhile. I would first like to thank the entire Materials Science department for even admitting such a long-shot candidate as me in the first place and for being an amazing group of teachers. In particular, I have had valuable conversations with professors Brent Fultz, Harry Atwater, and Julia Greer who have been very generous with their time. Harry was especially helpful with his input on the photovoltaic work as was his students Jeff Bosco and Greg Kimball. I appreciate that Julia kindly let me play with some cutting-edge fabrication work down in the cleanroom. Finally, I am very grateful to my advisor Axel who brought me to Caltech and has been supportive of my work these past years and has had to put up with the temperamental artist in me.

My group-mates Pratyush, Chirranjeevi (sp? - just kidding;), Ljuba, Qijun, Greg, and Ligen have made these past few years much more enjoyable than they would have been without their company. It has been said that a single conversation with a wise man is better than ten years of study and certainly my conversations with you (while not always about science or even particularly appropriate at times) have born that out. I am indebted for your support and friendship I hope we will walk along many of the same roads in the future. I am especially obliged to Ljuba for his help and many hours of work on the equation of state project, without whom it would never have been completed. Also Qijun was crucial in writing code to support the development of the perturbation-extrapolation method in the zinc phosphide project.

I have met so many fantastic people here which I hope to see in the future. In particular, I've been really fortunate to have met Cindy, Aron, Vanessa, Megan, Alex, Hanqing, and Christina. It has been a lot of fun and though I've seen more of the world than most, you're some of the most interesting, talented, and just plain nice people I've met. I'm famously bad about keeping in touch, but I hope you all put up with me and continue to at least poke me from time to time.

My parents who have always sacrificed so much. My brother and sister who provided a bit of a safety net for me to tight-rope over and who I should emulate more.

Thanks to my precious Bobbie for all the extra smiles this past year and in the years to come.

And finally to GG for always being there, my tower - none of this would have been possible without you.

Abstract

In this work we chiefly deal with two broad classes of problems in computational materials science, determining the doping mechanism in a semiconductor and developing an extreme condition equation of state. While solving certain aspects of these questions is well-trodden ground, both require extending the reach of existing methods to fully answer them. Here we choose to build upon the framework of density functional theory (DFT) which provides an efficient means to investigate a system from a quantum mechanics description.

Zinc Phosphide (Zn_3P_2) could be the basis for cheap and highly efficient solar cells. Its use in this regard is limited by the difficulty in n-type doping the material. In an effort to understand the mechanism behind this, the energetics and electronic structure of intrinsic point defects in zinc phosphide are studied using generalized Kohn-Sham theory and utilizing the Heyd, Scuseria, and Ernzerhof (HSE) hybrid functional for exchange and correlation. Novel 'perturbation extrapolation' is utilized to extend the use of the computationally expensive HSE functional to this large-scale defect system. According to calculations, the formation energy of charged phosphorus interstitial defects are very low in n-type Zn_3P_2 and act as 'electron sinks', nullifying the desired doping and lowering the fermi-level back towards the p-type regime. Going forward, this insight provides clues to fabricating useful zinc phosphide based devices. In addition, the methodology developed for this work can be applied to further doping studies in other systems.

Accurate determination of high pressure and temperature equations of state is fun-

damental in a variety of fields. However, it is often very difficult to cover a wide range of temperatures and pressures in an laboratory setting. Here we develop methods to determine a multi-phase equation of state for Ta through computation. The typical means of investigating thermodynamic properties is via 'classical' molecular dynamics where the atomic motion is calculated from Newtonian mechanics with the electronic effects abstracted away into an interatomic potential function. For our purposes, a 'first principles' approach such as DFT is useful as a classical potential is typically valid for only a portion of the phase diagram (i.e. whatever part it has been fit to). Furthermore, for extremes of temperature and pressure quantum effects become critical to accurately capture an equation of state and are very hard to capture in even complex model potentials. This requires extending the inherently zero temperature DFT to predict the finite temperature response of the system. Statistical modelling and thermodynamic integration is used to extend our results over all phases, as well as phase-coexistence regions which are at the limits of typical DFT validity. We deliver the most comprehensive and accurate equation of state that has been done for Ta. This work also lends insights that can be applied to further equation of state work in many other materials.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation for computer simulation in materials science

Throughout history mankind has probed the world around him with increasingly sophisticated tools. As technology progressed more aspects of the physical world were opened to exploration, which in turn drove technical developments further. Presently we have the ability to inspect the atomic structure of materials and observe phenomenon at femtosecond timescales. We can probe phonon modes via neutron scattering and watch dislocations form within nanostructures in real-time. With such capability, there is a question of the need for computer simulation at all.

The answer lies in the fact that the computer is a fundamentally new type of instrument which has become a means to imitate rather than directly observe the natural world. The great benefit here is that we have total control over this simulated 'world'. Conditions which are difficult and/or dangerous to achieve physically (such as extremes of temperature and pressure) can be simulated efficiently and safely. It is trivial to measure the melting point of a piece of iron at atmospheric pressure, it becomes a far harder task to measure it under the great pressures found in the Earth's core. Moreover, we can trivially test a huge number of very specific variations in materials, which may be expensive or impossible to explore experimentally.

Finally, even with the high resolutions of modern equipment, the mechanisms producing an experimental observation are often hard to determine as multiple physical processes are often convolved together to produce a certain effect. It is easy experimentally to determine a doping condition in a semiconductor, but it is often difficult to determine the microscopic cause of that state. In a simulation the entire history and all the relevant physics are open to scrutiny, which greatly simplifies determining causation.

The question then becomes how to trust our simulation as it necessarily always assumes some physical model for a system. The most obvious solution is not to rely on any abstraction at all and model the system from first principles - the so-called *ab initio* method. Our most fundamental view of the world relies on quantum mechanics so this would be the logical place to start formulating a description of the material world. Proceeding in this manner, it soon becomes apparent that a straightforward application of this framework to a physical problem is a hopelessly complex task for anything but trivial systems. A coarse numerical treatment of even simple molecules would require more datapoints than the number of particles in the universe - some level of abstraction will always be needed. Even though the much heralded Moore's Law highlights the exponential rate of increase in computing power since the 1970's, the main effect of this is to increase the size of a system that can practically be explored with a particular method rather than precluding any abstractions whatsoever (and this even ignores the fact that Moore's Law is fated to slow down in the near future). Increasing computing power, though it cannot be the whole answer, has allowed us to apply and verify more advanced techniques to real-world problems (as in the shift from Hartree-Fock to density function theory over the past 50 years). All the different methods in computational physics are essentially defined by what abstractions they make to the underlying quantum mechanics of a system. In using any of these procedures (or developing our own) we must always be aware of the assumptions we are making and develop useful abstractions that preserve the physics that are the most pertinent to the problems we are interested in. As we

shall see in the next section, the majority of the essential physics requires solving for the electronic structure.

1.2 Central problem in *ab initio* simulation

The fundamental description any physical system is that system's wavefunction. This is the maximally complete description of the subatomic, atomic, and even macroscopic properties a system possesses. The wavefunction is a solution to a linear partial differential equation - Schrödinger Equation (SE) - which describes how the quantum state of a system evolves with time:

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi \tag{1.1}$$

where Ψ is the wavefunction and $\hat{H}$ is the hamiltonian of the system. Generally, Ψ is a many-body function containing coordinates for all the degrees of freedom in a system, as well as time. As such, no closed form solutions exist for non-trivial cases; much like the n-body problem for gravity, analytic solutions to the SE are limited to '2-body' problems such as the hydrogen atom containing one electron and a nucleus.

Furthermore, numerical solutions are geometrically more difficult to solve as the complexity of a system increases. Just solving for the electrons in an oxygen atom requires solving a 24-dimensional PDE. Moving to even a microscopic bulk solid would be a hopelessly complex task. The trick will be to find approximations to the comprehensive wavefunction and SE treatment that will still retain enough accuracy for our purposes.

Towards this end, it is useful to make a distinction between physical properties that are determined from the lowest energy or ground electronic state and those that arise from excited states. In the vast majority of materials and conditions encountered commonly, the electronic states almost completely determine the physical properties.

Being much lighter than the nuclei, the electrons oscillate at a far higher velocity; since kinetic energy is linearly proportional to mass and proportional to the *square* of the velocity, electronic kinetic energy is substantially higher than ionic. Hence the energy ranges for their degrees of freedom are much higher. Another consequence is the time scale for nuclear motion is much longer than for electronic relaxation. As a result, the nuclei can be thought of as moving within a potential formed by the instantaneous electron ground state (suggesting the Born-Oppenheimer approximation discussed in section 2.2). Consequently, the structure and low-energy ionic motion of a solid (including properties such as crystal structure, cohesive energy, charge density, elastic constants, and nuclear vibration) will be determined almost entirely by the electronic ground state.

Since the electrons occupy a far larger volume than the nucleus due to their light mass and the wave nature of matter, and since their energy ranges are far larger than for the nuclei, electronic states also dominate how a material reacts to its environment. For a given arrangement of nuclei, the electronic excited states determine most transport and absorption phenomena such as dielectric response, electrical conductivity, and photoemission. However, these are often small disturbances of the entire system. Subsequently, perturbation theory can be used to build them up from a ground state description of the system. Thus, even for excited state properties, finding an efficient and accurate method to determine the electronic ground state of a system is central.

1.3 Overview

Determining the best approach for any particular problem is still in the realm of an art in many cases. In this work we chiefly deal with two broad classes of questions in computational physics, determining the doping mechanism in a semiconductor and developing an extreme conditions equation of state for a material. While solving certain aspects of these problems is well-trodden ground, both of them required ex-

tending the reach of existing methods to achieve their project's goals.

For the doping work we need to determine why a prospective photovoltaic material (zinc phosphide) cannot be n-type doped to produce a p-n homojunction. From the available experimental knowledge it seemed like the best place to look is at the intrinsic defects of the system. Each defect's formation energy and effect on the bandstructure needs to be meticulously calculated. Towards this end we develop a careful treatment of the thermodynamics as well as enumerate the various corrections to DFT needed for an accurate result. For a low-symmetry crystal like zinc phosphide this also required development of a new way to apply advanced functionals to large systems. The workflow for this project is presented in a general way which should be useful to study any semiconductor system. In the end, we are able to tell a convincing story about the defect relation to the doping and put forth suggestions to solve the issue in this particular case.

In developing the equation of state we need a robust way to calculate thermodynamic properties across a huge range of conditions. Molecular dynamics is the typical means to efficiently investigate thermodynamic properties. For the extreme conditions and accuracy needed for our purposes, we chose to use an *ab initio* approach to calculating the intermolecular forces - density functional theory. Even with such an approach we still need to widen the typical region of DFT validity. We detail how to extend DFT to finite temperatures as well as the statistical mechanics needed to capture as wide a region of phase-space as possible. Our results are compared to all available experimental and prior theoretical work and the final phase diagram is delivered as a table of free energies which can be used as input to various multi-scale methods in the future.

In the next chapter we start presenting a tractable framework for solving these types of problems. First, by approaching the solution of the SE through several approximations; this includes an in-depth discussion of DFT and its limitations, as well

as hints about possible workarounds. Afterwards, we discuss the specific projects and the task-specific extensions to DFT required to reach our goals.

Chapter 2

Theoretical background

2.1 Introduction

As discussed previously, the ground state electronic structure is of fundamental importance to practically all the physical properties of most systems. This chapter continues our consideration of the Schrödinger equation, in particular the approximations which can be used to accurately determine the electronic ground state of the system.

The ground state wavefunction can be found by minimizing the total energy with respect to all the degrees of freedom in Ψ subject to the constraints of normalization and particle symmetry. For an eigenstate, the time-independent value of the energy is the expectation value of the Hamiltonian:

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{2.1}$$

The eigenstates of the time-independent SE are stationary points of Eq.(1.1). Thus they can be found as solutions to the time-independent SE:

$$\hat{H}|\Psi\rangle = E|\Psi\rangle \tag{2.2}$$

Therefore, our starting point in finding a ground-state solution is the hamiltonian for a system of interacting electrons and nuclei:

$$\hat{H} = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \frac{\hbar^2}{2M_I}\sum_l \nabla_l^2 - \sum_{i,I}\frac{Z_I e^2}{|r_i - R_I|} + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|r_i - r_j|} + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J e^2}{|R_I - R_J|} \tag{2.3}$$

here the electrons are enumerated by lower case subscripts and the nuclei by upper-case. Electrons have a mass m_e, nuclei have mass M_I and charge Z_I, R_I and r_i are the positions of the ions and electrons respectively. This is the complete many-body hamiltonian including coulombic interactions between the electrons and between the electrons and the nuclei. It is useful to highlight the components of the hamiltonian for the coming discussions, typically this is done from the electronic point of view so we rewrite Eq(2.3) as:

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ext} + \hat{V}_{int} + \hat{V}_{n-n} \tag{2.4}$$

where T_e, T_n are the kinetic energy of the electrons and ions respectively. From the electronic POV, the potential of the nuclei are viewed as 'external' to the electrons, the potential of the electrons with themselves as 'internal'. As such, V_{ext} and V_{int} are the coulombic potential energy of the nuclei and electrons, and the electrons with themselves. The coulombic interaction of the nuclei with one another is accounted for in the final potential term $\hat{V}_{n-n}$.

As stated above, the Schrödinger equation with this hamiltonian cannot be solved analytically for more than two bodies (ie. the hydrogen atom). Even numerically, this hamiltonian leads to a system which is practically impossible to solve with reasonable accuracy for all but trivial systems. In the coming sections we detail the approximations used to turn this fully-interacting, multi-body SE into a tractable problem.

2.2 Born-Oppenheimer approximation

In order to solve for the wavefunction of even a single acetone molecule we would need to solve the time-independent Schrödinger equation in the spatial coordinates of the electrons and the nuclei (with 10 nuclei and 32 electrons this is a partial differential equation of 126 variables). Fortunately, in the many-body hamiltonian of Eq.(2.3) there is a large asymmetry in the kinetic energy of the electrons versus the ions due to the much larger mass of the ions ($M_I \gg m_e$). This results in very different time-scales for electronic and nuclear motion which would suggest a natural division of our complex problem into two simpler ones. In the Born-Oppenheimer (BO) approximation we break the total wavefunction of our system into its electronic and nuclear parts:

$$\Psi_{tot} = \Psi_{elec} \times \Psi_{nuc} \tag{2.5}$$

First, we assume that the ions are so massive (and thus move so slowly) that their motion can be ignored and we solve for the electronic part of the wavefunction assuming the nuclei are static (i.e. the electrons interact within a fixed nuclear potential). This involves solving an electronic SE:

$$\hat{H}_e(r, R)\Psi_{elec}(r, R) = E_e(R)\Psi_{elec}(r, R) \tag{2.6}$$

here r is the electronic coordinates and R the coordinates of the nuclei. The electronic hamiltonian ($\hat{H}_e$) does not include the kinetic energy of the nuclei $\hat{T}_n$. It is important to note that the electrons still interact with the nuclei through their respective coulombic potential energy terms. The electronic energy eigenvalues (E_e) are dependent on the ionic positions (R) which enter as parameters. Once Ψ_{elec} is solved for we then use it to generate the potential for the SE involving only the nuclei:

$$[\hat{T}_n + E_e(R)]\Psi_{nuc}(R) = E\Psi_{nuc}(R) \tag{2.7}$$

where E is the energy eigenvalues for the entire system. Again, because the nuclei are far more massive than the electrons, the nuclear positions (R) are updated in small increments and the electronic wavefunctions are completely recomputed for each step. This is reminiscent of the adiabatic theorem so the BO approximation is often referred to as the adiabatic approximation. For the example of an acetone molecule, the solution is arrived at in two steps, first the electronic SE is solved with 96 electronic coordinates. Then the nuclear SE is solved with an equation of 30 variables. In many instances, the nuclear kinetic and interaction potential V_{n-n} are just solved classically and added to the electronic result which would make the solution of the nuclear SE unnecessary.

The BO approximation is an important computational speedup and is generally an excellent estimation for most purposes. Even if electron-phonon interactions are of critical importance (where we cannot strictly separate the electronic and nuclear coordinates) the BO approximation is still a good starting point for solving these systems via a perturbative approach.

2.3 Pseudopotentials

In most common phenomena the great majority of the physics only involves valence electrons. The idea of a pseudopotential is to replace the strong Coulomb potential of the ions and tightly bound core electrons with a simpler 'effective potential' that acts on the valence electrons while preserving the pertinent physics.

When an electron scatters off a localized spherical potential, its wavefunction undergoes an energy-dependent phase shift. The central idea of a pseudopotential is that all the properties of a wavefunction outside of the scattering region is completely determined by this phase shift and is invariant to a shift by any multiple of 2π. Thus, we can replace the actual potential with one which has more desirable characteris-

tics, and still preserve the properties of the valence wavefunctions outside of the core region of the ion. Developing a pseudopotential consists of finding functions which reproduce the scattering phase-shift (modulo 2π) over a desired energy range.

A great advance in the development of the pseudopotentials in common use today are the so-called 'norm-conserving' pseudopotentials [1]. Norm-conserving pseudopotentials are defined by the following properties [2]:

1. all-electron (AE) and pseudo-valence eigenvalues agree for a chosen 'prototype' atomic configuration
2. AE and pseudo-valence wavefunctions agree beyond a chosen 'core radius' R_c
3. the integrals for $0<r<R_c$ for the all-electron and pseudo charge densities agree (norm-conservation)
4. the logarithmic derivatives of the all-electron and pseudo wavefunctions and their *first energy derivatives* agree for $r \geq R_c$

Properties (1) and (2) ensures that the pseudopotential is equivalent to the atomic potential outside the 'core region' of radius R_c. This is a direct result of both potentials being solutions to the same differential equation and homogeneous boundary condition for $r>R_c$.

Property (3) requires that the integrated charge:

$$Q(R_c) = \int_0^{R_c} dr r^2 |\Psi(r)|^2 \tag{2.8}$$

is the same for the AE and pseudo wavefunctions for all valence states. This ensures that the the total charge in the core region is correct and also that the pseudo wavefunction is equal to the true orbital for $r>R_c$. As a consequence, the potential outside of R_c is correct as well since the potential outside a spherically symmetric charge distribution depends only on the charge enclosed by the sphere.

Point (4) is actually a direct result of the norm-conserving property (3). This can be shown by first writing the radial equation for a spherical potential:

$$-\frac{1}{2}\phi''(r) + \left[\frac{l(l+1)}{2r^2} + V(r) - \epsilon\right]\phi(r) = 0 \tag{2.9}$$

where $\phi(r)$ is a single particle wavefunction. If we transform this equation by defining a variable $x(\epsilon, r)$:

$$x(\epsilon, r) \equiv \frac{d}{dr} ln\phi(r) \tag{2.10}$$

we can show that Eq.(2.9) is equivalent to:

$$x(\epsilon, r)' + x(\epsilon, r)^2 = \frac{l(l+1)}{r^2} + 2\left[V(r) - \epsilon\right] \tag{2.11}$$

taking the general form of this equation and differentiating w.r.t. energy gives:

$$\frac{\partial}{\partial\epsilon}x(\epsilon, r)' + 2x(\epsilon, r)'\frac{\partial}{\partial\epsilon}x(\epsilon, r) = \frac{1}{\phi(r)^2}\frac{\partial}{\partial r}\left[\phi(r)^2\frac{\partial}{\partial\epsilon}x(\epsilon, r)\right] \tag{2.12}$$

multiplying by $\phi(r)^2$ and integrating, at radius R we find expressions for the derivative of the energy and the logarithmic energy derivative, $D(\epsilon, r)$, respectively:

$$\frac{\partial}{\partial\epsilon}x(\epsilon, R) = -\frac{1}{\phi(R)^2}Q(R) \tag{2.13}$$

$$\frac{\partial}{\partial\epsilon}D(\epsilon, R) = -\frac{R}{\phi(R)^2}Q(R) \tag{2.14}$$

by inspection, it is clear that if the pseudo wavefunction and the AE wavefunction have the same magnitude at R_c and the norm-conserving property holds ($Q_{pseudo}(R)$ = $Q_{AE}(R)$), then both the first energy derivative and the logarithmic derivative of the pseudo wavefunction is identical to the AE wavefunction. This ensures that the pseudopotential has the same scattering phase as the all-electron atom to linear order near a specified energy.

This last property is crucial for the pseudopotential to have *transferability* between a variety of chemical environments. That is, we can develop the pseudopotential in a simple environment and as long as property (4) is satisfied, use it with great accuracy in a more complex system.

The benefits of using pseudopotentials are two-fold. First, there are fewer electrons to solve the SE for as we assume that the core electrons are unaffected by chemical bonding. Secondly, the valence electronic wavefunctions do not have to be represented close to the ionic cores where they oscillate strongly and have many nodes. This is desireable for reducing the size of the basis set needed to accurately represent the wavefunctions as they require less frequency content in their description. As we shall see, a convenient choice of basis for solid-state systems are plane-waves; pseudopotentials reduce the number of higher frequency plane-waves that need to be considered.

2.4 Density functional theory

Even with the BO approximation and the use of pseudopotentials, the Schrödinger Equation, with the multi-body hamiltonian of Eq.(2.3) is still much too complex to solve generally. Density functional theory (DFT) is a step on the way to a reformulation the problem from many interacting particles to independent particles acting within an effective potential. The effective potential being a mean-field resulting from the other particles in the system.

The principal idea is that all the properties of the system can be thought of as a function of the ground state density $n_0(r)$ (i.e. any property is a *functional* of the density). To put it another way, knowing the ground state density is equivalent to knowing the wavefunction for the ground state *and all excited states* as well. As we build to the Kohn-Sham formalism we no longer require the hamiltonian of Eq.(2.3) operating on a 3N-dimensional wavefunction. We can work with independent-particle

wavefunctions and can derive all the attributes of our system from scalar functions of $n_0(r)$ in three dimensions.

As suggested by the BO approximation, it is typical to consider the electrons separately from the nuclei. The nuclei are often abstracted away as an external potential V_{ext} in which the electrons relax into their ground state. The ionic kinetic and self-interaction potential energy are considered separately (often as a classical calculation). The hamiltonian of Eq.(2.3) for the electrons becomes:

$$\hat{H} = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 + \sum_i v_{ext}(r_i) + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|r_i - r_j|} \tag{2.15}$$

This complex multi-body problem cannot be broken up into simpler single-body equations due to the electron coulombic self-interation term. The Hohenberg-Kohn Theorems provide a first step into a solution to this problem by assigning a special significance to the ground state particle density.

2.4.1 Hohenberg-Kohn theorems

DFT is predicated on two theorems first put forth by Hohenberg and Kohn:

Theorem 2.1: *In any system of interacting particles, the external potential $v_{ext}(r)$ is uniquely determined (up to a constant) by the ground state particle density $n_0(r)$.*

Theorem 2.2: *For any $v_{ext}(r)$ we can define a functional for the energy in terms of particle density E[n(r)]. The density that minimizes this functional is the correct ground state density $n_0(r)$.*

The proof of theorem 2.1 is best shown by assuming two external potentials $v_{ext}^{(1)}(r)$ and $v_{ext}^{(2)}(r)$ that differ by more than a constant and both give rise to the same charge density n(r). These two different potentials give rise to two different hamiltonians ($\hat{H}^{(1)}$ and $\hat{H}^{(2)}$) and two different ground state wavefunctions ($\Psi^{(1)}$ and $\Psi^{(2)}$). Since $\Psi^{(2)}$ is not the ground state of $\hat{H}^{(1)}$ we know:

$$E^{(1)} = \langle \Psi^{(1)} | \hat{H}^{(1)} | \Psi^{(1)} \rangle < \langle \Psi^{(2)} | \hat{H}^{(1)} | \Psi^{(2)} \rangle \tag{2.16}$$

For simplicity we consider a non-degenerate system, though the proof can be readily extended to degenerate systems as well [3]. The last term in Eq.(2.16) can be expressed as:

$$\langle \Psi^{(2)} | \hat{H}^{(1)} | \Psi^{(2)} \rangle = \langle \Psi^{(2)} | \hat{H}^{(2)} | \Psi^{(2)} \rangle + \langle \Psi^{(2)} | \hat{H}^{(1)} - \hat{H}^{(2)} | \Psi^{(2)} \rangle \tag{2.17}$$

$$= E^{(2)} + \int d^3r [v_{ext}^{(1)}(r) - v_{ext}^{(2)}(r)] n_0(r) \tag{2.18}$$

and from the initial inequality:

$$E^{(1)} < E^{(2)} + \int d^3r [v_{ext}^{(1)}(r) - v_{ext}^{(2)}(r)] n_0(r) \tag{2.19}$$

However, if we start from Eq.(2.16) and consider $E^{(2)}$ in a similar fashion we obtain the same result with interchanged superscripts:

$$E^{(2)} < E^{(1)} + \int d^3r [v_{ext}^{(2)}(r) - v_{ext}^{(1)}(r)] n_0(r) \tag{2.20}$$

by adding 2.19 and 2.21 we obtain a contradiction:

$$E^{(1)} + E^{(2)} < E^{(1)} + E^{(2)} \tag{2.21}$$

so our initial assumption that two external potentials differing by more than a constant can give rise to the same charge density is proven false. Thus, the ground state particle density defines the external potential up to a constant.

A corollary to this is that since the hamiltonian is uniquely determined (up to a constant) then the wavefunction for all states (*excited* as well as ground) are also determined; thus all the properties of the system can, in principle, be derived from only the ground state density, although this mapping may be too complex to be of

practical use.

The proof of theorem 2.2 relies on the previous corollary, that is, all the properties of a system are uniquely determined by the particle density $n(r)$. Every property (including the total energy E_{HK}) can be expressed as a *functional* of $n(r)$ and we can rewrite Eq.(2.4) as:

$$E_{HK}[n] = T_e[n] + E_{int}[n] + \int d^3r v_{ext}(r)n(r) = F_{HK}[n] + \int d^3r v_{ext}(r)n(r) \quad (2.22)$$

here we are just considering the total energy of the electrons and ignore the nuclear kinetic and nuclei-nuclei Coulombic potential energy to simplify the discussion. Strictly speaking, this proof is only valid for densities $n(r)$ that can be generated from some potential v_{ext} or "V-representable." An alternate functional definition due to Levy and Lieb loosens this restriction to densities that can be generated from an N-electron wavefunction or "N-representable"; such wavefunctions can be found for any density satisfying simple conditions. Thus functionals can be found for any density which satisfy these undemanding limitations [4].

The density-dependant functional $F_{HK}[n]$ encapsulates the kinetic T_e and potential energy E_{int} of the interacting electrons. Consider that any trial density $\tilde{n}$ defines its own hamiltonian and hence its own wavefunction $\tilde{\Psi}$. If $\tilde{\Psi}$ is used as a trial wavefunction for the hamiltonian generated by the external potential v_{ext} corresponding to the ground state density $n_0(r)$ and wavefunction $\Psi^{(0)}$ we have:

$$E[\tilde{n}] = \langle\tilde{\Psi}|\hat{H}|\tilde{\Psi}\rangle > \langle\Psi^{(0)}|\hat{H}|\Psi^{(0)}\rangle = E[n_0] \quad (2.23)$$

The energy functional E[n] always gives the minimum energy for the ground state density. Furthermore, if the functional $F_{HK}[n]$ is known, we can vary the particle density n[r] until we minimize the total energy of the system. Thus knowing E[n] is sufficient to determine the exact ground state density and energy. The key problem

here though is that the functional $F_{HK}[n]$ relies on many-body correlations and is not typically known for anything but the simplest systems.

In theory, we can extract any property from the ground state density of our system. However, generally the relations from density to physical properties tend to be very subtle even for simple systems. Considering the case of non-interacting electrons in an external potential; while the exact functional $F_{HK}[n]$ is simply the kinetic energy, there is no known way to proceed directly from the density and obtain the kinetic energy of the system. One has to determine the set of N multi-body wavefunctions and proceed from there. As such, DFT would be of limited value if it were not for the further refinement of the Kohn-Sham approach that allows us to make approximations to the real ground state functionals and still retain the essential physics.

2.4.2 Kohn-Sham equations

The approach suggested by Kohn and Sham is to reformulate the multi-body hamiltonian of Eq.(2.3) as an independent particle system *which produces exactly the same ground state density*. In essence, the electrons are assumed to be uncorrelated except for what is required to satisfy the exclusion principle. An effective potential, that is purely a function of particle density, is then included to approximate the effects of many-body correlation, but there is no interaction term explicitly in the hamiltonian. Our problem has now become one where the *densities* interact rather than the particle wavefunctions.

The critical fact here is that it can be shown that the ground state density of the interacting many-body system can be generated from a simpler system of independent particles. All the multi-body interactions are abstracted away into an effective potential in which the electrons move. We are left with a set of independent particle SE's which are exactly solvable (in theory):

$$(-\frac{\hbar^2}{2m}\nabla^2 + v_{eff})\phi_i = \epsilon_i\phi_i \tag{2.24}$$

where ϕ_i are the independent particle orbitals. For a system with N electrons, the ground state would consist of the N lowest energy orbitals. We obtain the density of the system by summing the contributions of each orbital:

$$n(r) = \sum_i^N |\phi_i(r)|^2 \tag{2.25}$$

We now reformulate the many-body HK energy functional Eq.(2.22) in the following independent-particle form:

$$E_{KS}[n] = T_s[n] + \int d^3r v_{ext}(r)n(r) + E_H[n] + E_{XC}[n] \tag{2.26}$$

where we now have $T_s[n]$ as the independent particle kinetic energy:

$$T_s[n] = \sum_{i=1}^N \langle\phi_i| - \frac{\hbar}{2m}\nabla^2|\phi_i\rangle \tag{2.27}$$

and we have broken the multi-body electron-electron interaction energy E_{int} into a classical coulomb self interaction term, or Hartree energy $E_H[n]$:

$$E_H[n] = \frac{1}{2}\int d^3r d^3r' \frac{n(r)n(r')}{|r-r'|} \tag{2.28}$$

and a term that accounts for all the many-body quantum mechanical effects of exchange and correlation E_{XC}. By equating the energy functionals in Eq.(2.26) and Eq.(2.22) we can rewrite this as:

$$E_{XC}[n] = (T_e[n] - T_s[n]) + (E_{int}[n] - E_H[n]) \tag{2.29}$$

which highlights that this is the residue kinetic and potential energy in approximating the fully-interacting many-body system as a system of independent-particles. This energy (and its associated potential) are the only unknowns in the KS approach.

While not essential, the great advantage of recasting the HK energy functional in the KS form is that in separating the independent-particle kinetic energy and the long-range Hartree terms, the remaining contribution from $E_{XC}[n]$ can be reasonably approximated as a local or semi-local functional. The success of the KS method is that many systems can be investigated accurately with even gross approximations to this term.

2.4.2.1 Variational expression and self-consistent solution

Solving the KS equations requires solving the independent particle SE Eq.(2.24) under the constraint that the effective potential v_{eff} and density n(r) are consistent. In practice, this is done numerically by successively changing n and v_{eff} to achieve self-consistency as shown in Fig.(2.1). First, an initial guess is made about the electron density. The effective potential is then calculated by considering the variation of the total energy functional with respect to electron density. This yields an effective potential:

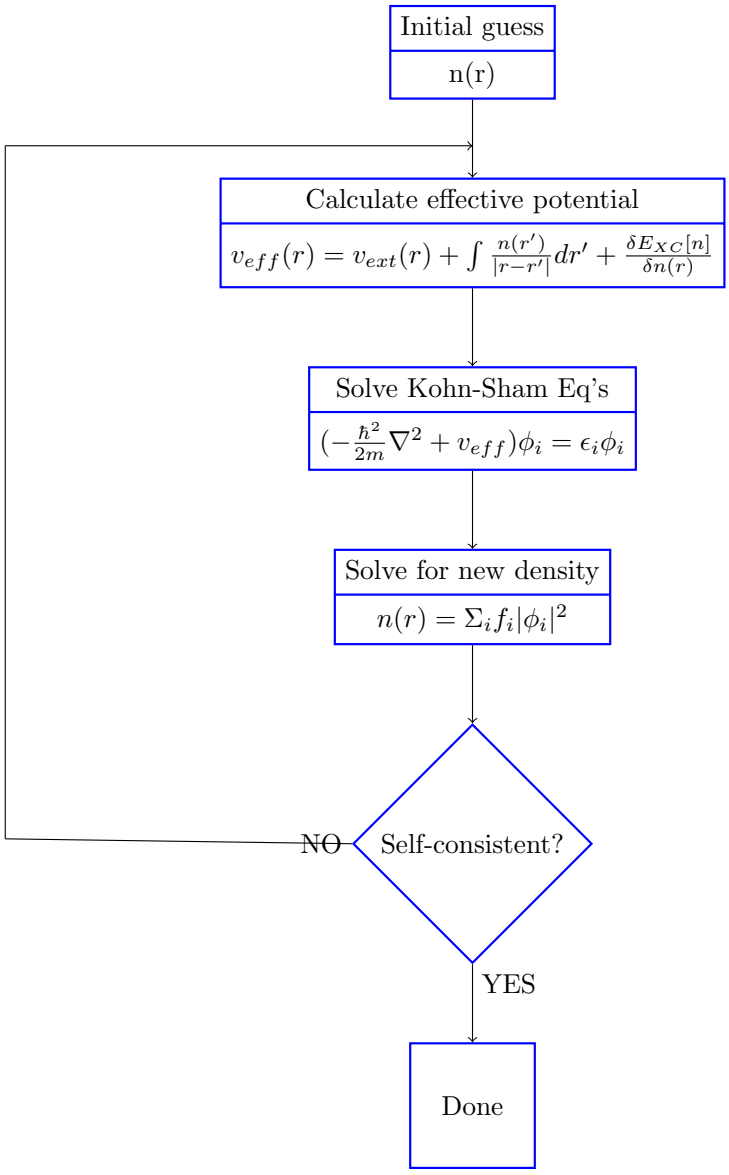
$$v_{eff}(r) = v_{ext}(r) + \int \frac{n(r')}{|r - r'|} dr' + \frac{\delta E_{XC}[n]}{\delta n(r)} \tag{2.30}$$

With this effective potential, we solve Eq.(2.24). This is an eigenvalue problem and is where most of the computational effort is spent, so much effort is spent to optimize this step. Finally, we calculate the new particle density from Eq.(2.25) - due to stability issues, the new density is usually some average of the old and new densities - and check for convergence. If we are self-consistent we have a solution, if not then we recalculate v_{eff} for the new density and then run the inner loop again.

2.4.2.2 Meaning of KS eigenvalues

Working within the KS framework, we depict a multi-body system as a set of single particle orbitals ϕ_i. In theory, they should not have a precise physical meaning, as

Figure 2.1: Flowchart diagramming the self-consistent loop used in solving the KS equations.



we are trying to describe a multi-body system with a collection of independent particles. In practice, though, it has been observed that the eigenvalues give a reasonable representation of the band structure of many materials.

Within KS formalism, the eigenvalues have a precise *mathematical* meaning. In his continuous N extension of KS theory, Janak proved that:

$$\frac{\delta E}{\delta n_i} = \epsilon_i \tag{2.31}$$

with n_i being the occupation number of the ith orbital [5]. As for a physical interpretation, the exact KS theory only makes guarantees about the highest occupied state's eigenvalue. This can be understood by considering that the true multi-body fully interacting system must have the same ionization energy as the KS system since they are both assumed to have the same density (see corollary to Thm. 2.1 above). In addition, as one electron coordinate of our non-interacting ground-state N-electron wavefunction tends to infinity, it can be shown that the remaining electrons collapse to the (N-1) electron ground-state; thus, the electron density of the N-electron system decays exponentially with this electron coordinate [6]. This requires that the KS potential is zero at infinity leading to the result that the ionization energy is just the negative of the energy of the highest occupied orbital's eigenvalue. Of course, approximations to the exchange-correlation potential can lead to significant errors for the calculated ionization energy.

While only the highest occupied eigenvalue has a direct physical interpretation, Janak's Theorem is often 'bent' to give a physical meaning to the other eigenvalues as being electron removal/addition energies from those states. Indeed, a good approximation to the band structure is often obtained in this manner. This relies on the assumption that the addition or subtraction of an electron from a state does not change the potential of the other orbitals drastically. Consequently, extended states are often better described than more localized states as the the removal of a highly

localized electron will have a large effect on its neighbors.

The sum of the KS eigenvalues also have a physical meaning which can be derived from multiplying the KS equation (2.24) by $\phi^*(r)$, integrating over space, and summing over i. This leads to a relation:

$$\sum_i \epsilon_i = T_s[n] + \int v_{eff}(r)n(r)dr \tag{2.32}$$

where the sum of the eigenvalues is equal to the kinetic and potential energy of the independent-particle system. From Eq.(2.30) and (2.26) the total energy can then be expressed as:

$$E_{KS}[n] = \sum_i^N \epsilon_i - E_H[n] + E_{XC}[n] - \int \frac{\delta E_{XC}[n]}{\delta n(r)} n(r)dr \tag{2.33}$$

As we shall see when we discuss the exchange-correlation potential, we must always exercise care when comparing KS eigenvalues to physically meaningful traits.

2.4.3 Independent-particle problem: exchange and correlation

In order to reframe our problem from a many-body, fully interacting system to one in terms of independent-particles we have to abstract away all the many-body coupling terms into the exchange and correlation energy E_{XC}. In Eq.(2.29) we have defined E_{XC} as the energy difference between the exact, fully-interacting system and independent particles within a purely classical Hartree potential. This difference can be understood in terms of two effects due to quantum mechanics, exchange and correlation.

Exchange accounts for Pauli exclusion and corrects for the self interaction inherent in the Hartree energy (i.e. an electron spuriously interacts with its own density). Exchange interactions tend to separate electrons with the same spin, thus lowering the total energy by reducing their Coulombic repulsion. These effects can be calcu-

lated exactly for some independent-particle approaches such as Hartee-Fock, but not generally for KS methods.

Correlation effects are the result of the collective behavior of electrons to screen each other and reduce their overall coulombic interaction. These interactions are more pronounced for electrons with opposite spins as they are more likely to occupy nearby locations. Correlation is the residue of reducing the many-body wavefunction to an independent-particle form and typically cannot be calculated analytically, even in most trivial systems.

Since both exchange and correlation tend to keep electrons apart, these effects can be described in terms of a 'hole' surrounding each electron which prevents other electrons from approaching it. This *exchange-correlation hole* can be understood in terms of the joint probability of finding an electron at point r given that one exists at point r':

$$\begin{aligned} n(r,\sigma,|r'\sigma') &= \left\langle \sum_{i\neq j} \delta(r-r_i)\delta(\sigma-\sigma_i)\delta(r'-r_j)\delta(\sigma'-\sigma_j) \right\rangle \\ &= N(N-1)\sum_{\sigma_3,\sigma_4,\ldots}\int dr_3 \ldots dr_N |\Psi(r,\sigma;r',\sigma';r_3,\sigma_3;\ldots r_N,\sigma_N)|^2 \end{aligned} \tag{2.34}$$

A measure of the correlation $\Delta n(r,\sigma|r',\sigma')$ would be the difference between this and the independent particle probabilities:

$$\Delta n(r,\sigma|r',\sigma') = n(r,\sigma|r',\sigma') - n(r,\sigma)n(r',\sigma') = h_{XC}(r) \tag{2.35}$$

This is the exchange-correlation hole h_{XC} which is a concept we shall return to below.

2.4.3.1 Exchange correlation energy

We can separate the exchange-correlation energy E_{XC} from Eq.(2.29) into a kinetic and potential energy term:

$$E_{XC}[n] = (T_e[n] - T_s[n]) + (E_{int}[n] - E_H[n]) = T_C[n] + V_{XC}[n] \tag{2.36}$$

$$T_C[n] = T_e[n] - T_s[n] \tag{2.37}$$

$$V_{XC}[n] = E_{int}[n] - E_H[n] \tag{2.38}$$

These equations emphasize that E_{XC} is simply the sum of two effects: (i) the kinetic energy difference between a non-interacting, independent-particle system and the true multi body system due to correlation and (ii) the difference between the true internal potential of an Coulombic, many body system and an independent particle approximation in which the internal interactions have been replaced by the classical self-interaction energy of a charge distribution E_H.

2.4.3.2 Exchange correlation hole

To get a deeper level of understanding we can examine the exchange-correlation potential energy V_{XC}. As previously stated, it can be thought of as the difference in internal potential energy between a system described as a many-body wavefunction and the independent particle Hartree energy obtained via a double integral of density as in Eq.(2.28). The Hartree energy can be expressed in terms of a potential v_H:

$$E_H[n] = \int dr v_H(r) n(r) \tag{2.39}$$

where v_H is the Hartree energy per electron given by:

$$v_H(r) = \frac{1}{2}\int dr' \frac{n(r')}{|r - r'|} \tag{2.40}$$

An electron found at position r, within a charge distribution, has a potential energy of $v_H(r)$ due to its interaction with that charge distribution. Similarly, in the fully interacting system we can define v_{int} via:

$$E_{int}[n] = \int dr v_{int}(r) n(r) \tag{2.41}$$

where v_{int} is the internal potential energy per electron. Here matters are complicated in that we are trying to approximate a many-body wavefunction. Since the forces between particles in the many-body hamiltonian are two-body, pairwise correlation functions should be a good measure of many properties including this potential energy. Intuitively, this correlation means having an electron at point r affects the chance of having one at point r'.

The internal potential v_{int} at a point r is dependent on a conditional charge distribution $n(r', \sigma'|r, \sigma)$:

$$v_{int}(r) = \frac{1}{2}\int dr' \frac{n(r', \sigma'|r, \sigma)}{|r - r'|} \tag{2.42}$$

through Eq.(2.38) we arrive at:

$$v_{XC}(r) = \frac{1}{2}\int dr' \frac{n(r', \sigma'|r, \sigma) - n(r')}{|r - r'|} \tag{2.43}$$

where the numerator can be termed the *exchange-correlation hole*:

$$h_{XC}(r) = n(r', \sigma'|r, \sigma) - n(r') \tag{2.44}$$

which is reduction in the classical value of the electron density at r due to having a second electron at point r'. Every electron is surrounded by such an exchange-correlation hole, reducing its potential energy. Furthermore, this 'electron depletion

region' can be thought of the sum of two 'holes'; one resulting from purely exchange h_X, and one from only electron correlation effects h_C.

The exchange hole h_X can be thought of as a 'positive charge hole' that surrounds each electron and represents two effects, Pauli exclusion and the necessary corrections to the Hartree energy to cancel the false self-interaction energy (i.e. the electron interacting with its own density). It can be defined in terms of an independent particle system:

$$h_X(r'|r) = n_s(r', \sigma'|r, \sigma) - n(r') \tag{2.45}$$

where $n_s(r', \sigma'|r, \sigma)$ is the joint probability density of the non-interacting system. For a set of single-particle spin orbitals ϕ_i, this can be shown to be:

$$h_X(r'|r) = -\delta_{\sigma,\sigma'} \frac{|\sum_i \phi_i^{\sigma *}(r)\phi_i^{\sigma'}(r')|^2}{n(r)} \tag{2.46}$$

which is in fact the exact exchange as calculated in the Hartree-Fock method. This is a non-local operator so it is not used in most common KS approximations to E_{XC}, though as we shall see later, this is useful in so-called 'hybrid' functionals.

The correlation hole chiefly accounts for the collective motions of the electrons to screen themselves and is typically much more difficult to develop accurate approximations for. Towards further understanding of this effect, it is useful to consider the true system as an average of two extreme cases of electron coupling; an independent-particle system with no coupling, and a system of particles fully coupled via pairwise coulombic interaction.

2.4.3.3 Adiabatic connection

Up until now, the non-interacting and fully correlated many body systems have been thought of as entirely separate except for the fact that they have the same density. However, it is also possible to view them as two instances of a continuous set of systems

defined by the degree of electron-electron coupling λ. Returning to the many-body hamiltonian of Eq.(2.15), if we replace the potential energy operator with:

$$\hat{V}_\lambda[n] = \frac{1}{2}\sum_{i\neq j}\frac{\lambda}{|r_i - r_j|} + \sum_i V'_{ext}[n](r_i) \tag{2.47}$$

where $V'_{ext}[n](r_i)$ is some fictitious potential defined so as to keep the ground state density constant for all values of λ. A fully interacting system is given when $\lambda = 1$, the non-interacting is given by $\lambda = 0$, the continuous path between the two is the *adiabatic connection*. The ground state wavefunction minimizes total energy over all degrees of freedom (DOF) which are not constrained by normalization or anti-symmetry. Thus, a small change in any wavefunction DOF, at a constant λ, will not produce any first-order change in energy (a result of the "2n+1" theorem) so long as the constraints are maintained. Considering a small perturbation of the correlation hole $\delta h_C(r'|r)$ that keeps the density constant at a specific λ; the total energy remains constant and, as a result of the density also being fixed, we know the external, Hartree, non-interacting kinetic, and exchange energy is constant as well. From Eqs.(2.26) and (2.36):

$$\Delta E_{XC}[n] = 0 = \Delta E_X + \Delta E_C \tag{2.48}$$

$$\Delta E_C = \Delta V_C + \Delta T_C = 0 \tag{2.49}$$

$$\Delta T_C = -\Delta V_C \tag{2.50}$$

Now considering a small step along the adiabatic connection $\delta\lambda$, this produces a change in the correlation potential energy per electron δv_C due to both an increase in the Coulomb potential as well as the shape of the correlation hole itself:

$$\delta v_C(r) = \frac{d\lambda}{2}\int dr' \frac{h_C^{(\lambda)}(r'|r)}{|r - r'|} + \frac{\lambda}{2}\int dr' \frac{\delta h_C^{(\lambda)}(r'|r)}{|r - r'|} \tag{2.51}$$

where the second term is the change in correlation potential energy at a fixed λ. Referencing Eq.(2.50) we arrive at an expression for the change in total correlation energy per electron $\delta E_C(r)$:

$$\delta v_C(r) = \frac{d\lambda}{2}\int dr' \frac{h_C^{(\lambda)}(r'|r)}{|r-r'|} - \delta t_C(r) \Longrightarrow \delta E_C(r) = \frac{d\lambda}{2}\int dr' \frac{h_C^{(\lambda)}(r'|r)}{|r-r'|} \tag{2.52}$$

Integrating $\delta E_C(r)$ for $\lambda = 0$ to 1 gives us:

$$E_C(r) = \int_0^1 d\lambda \frac{1}{2}\int dr' \frac{h_C^{(\lambda)}(r'|r)}{|r-r'|} = \frac{1}{2}\int dr' \frac{\bar{h}_C(r'|r)}{|r-r'|} \tag{2.53}$$

where $\bar{h}_C(r'|r)$ is the *coupling constant averaged* correlation hole which defines the total correlation energy per electron rather than just the potential portion. Furthermore, since the exchange hole does not depend on λ we can define the *coupling constant averaged* exchange-correlation hole $\bar{h}_{XC}(r'|r)$:

$$\bar{h}_{XC}(r'|r) = \int_0^1 d\lambda \frac{h_{XC}^{(\lambda)}(r'|r)}{|r-r'|} \tag{2.54}$$

and we arrive at the total exchange-correlation energy per electron as:

$$\epsilon_{XC}(r) = \frac{1}{2}\int dr' \frac{\bar{h}_{XC}(r'|r)}{|r-r'|} \tag{2.55}$$

$\epsilon_{XC}[n]$ can be thought of as an interpolation of the exchange-only and fully correlated energies at a specific density n(r). E_{XC} can now be expressed as:

$$E_{XC}[n] = \int dr n(r)\epsilon_{XC}([n], r) \tag{2.56}$$

where $\epsilon_{XC}([n], r)$ is the exchange-correlation energy per electron at point r that depends on the local density n.

This derivation has produced some significant results, the most important being that if we know the shape of the exchange-correlation hole we can solve the many

body problem exactly. Also we have transformed the correlation kinetic energy T_C into a potential form that is related to the correlation hole $h_C(r'|r)$. This points to a link between the electron density and the shape of the many-body wavefunction by means of the exchange-correlation hole.

2.4.3.4 Exchange-correlation potential and the bandgap discontinuity

The exchange-correlation potential is defined as the functional derivative of E_{XC} with respect to electron density:

$$v_{XC}(r) = \epsilon_{XC}([n], r) + n(r)\frac{\delta\epsilon_{XC}([n], r)}{\delta n(r)} \tag{2.57}$$

where $\epsilon_{XC}[n]$ is defined as in Eq.(2.55). Here the second term accounts for the change in the exchange-correlation hole with density (i.e. a "response" function). In non-metals, the nature of the states changes discontinuously at the bandgap, leading to a discontinuity in the response term w.r.t. electron density. As a consequence, the addition of one electron changes the potential for every electron in the system.

According to Janak's Theorem, the KS eigenvalues are the change in the total energy with respect to state occupation:

$$\epsilon_i = \frac{dE_{tot}}{dn_i} = \int dr \frac{dE_{tot}}{dn(r)}\frac{dn(r)}{dn_i} \tag{2.58}$$

For a non-interacting system it is straightforward to calculate the eigenvalues directly. However, for a KS system, the derivative of the potential energy part of $\frac{dE_{tot}}{dn(r)}$ contains v_{XC}. As stated above, this has a response part which can change discontinuously with state occupation, which gives rise to spurious jumps in the eigenvalues. This is the well-known 'bandgap discontinuity' problem within DFT; that is, even in the exact KS theory, there is no guarantee that the difference between the highest occupied and lowest unoccupied states will equal the true bandgap.

The fundamental bandgap ϵ_{gap} can be identified as the second order difference of

the ground state energy with respect to electron number:

$$\epsilon_{gap} = E(N+1) + E(N-1) - 2E(N) \tag{2.59}$$

which in the KS framework can be decomposed into two contributions:

$$\epsilon_{gap} = \varepsilon_{gap} + \Delta_{xc} \tag{2.60}$$

where ε_{gap} is the difference between the energies of the highest occupied and lowest unoccupied states, and Δ_{xc} is the N-derivative discontinuity of the exchange-correlation functional $\Delta_{xc} = \frac{\delta\epsilon_{XC}([n],r)}{\delta n(r)}|_{N+\delta} - \frac{\delta\epsilon_{XC}([n],r)}{\delta n(r)}|_{N-\delta}$. It has been shown the Δ_{xc} that corresponds to the exact E_{XC} functional must be finite [7, 8]. However, all approximate functionals based only on electron density are continuous functions of N (i.e. $\Delta_{xc} = 0$).

Within KS theory there is no easy solution to this dilemma. One approach is to explore functional-based strategies such as the optimized effective potential (OEP) and hybrid functional methods (see Sec. 2.4.4.3) where orbital-dependent and non-local terms are used to help generate more accurate bandgaps. Another strategy is to use the KS eigenvalues as starting points in a perturbative method such as GW (see Sec. 2.5) in which the all the eigenvalues have a precise physical meaning.

2.4.4 Exchange correlation functionals

In the KS approach we consider the independent-particle kinetic energy and classical Coulombic interaction separate from the quantum effects of exchange and correlation. The central task in the KS technique is determining the exchange-correlation functional that accurately reproduces these effects. A great benefit of this approach is that since these quantum influences are typically short-ranged they can be well approximated as a local or semi-local functional of the electron density. In this section, some general forms of particular interest to solid-state systems are presented.

2.4.4.1 Local density approximation (LDA)

This class of functionals approximate exchange-correlation effects as solely depending on the electron density at a point of interest:

$$E_{xc}^{LDA}[n] = \int n(r)\epsilon_{xc}(n(r))dr \tag{2.61}$$

There are many different LDA-based approximations to the XC energy, but by far the most common is derived from assuming the solid approaches the homogeneous electron gas (HEG) limit. As before, we can decompose E_{xc}^{LDA} into separate contributions accounting for exchange and correlation. For a HEG, the exchange energy has a simple analytic form:

$$E_{x}^{LDA}[n] = -\frac{3}{4}(\frac{3}{\pi})^{\frac{1}{3}} \int n(r)^{\frac{4}{3}}dr \tag{2.62}$$

The correlation energy is more complex and is not known except in the low and high density limits. For intermediate densities, the typical practice is to fit quantum Monte Carlo simulation data for the correlation energy.

As an approximation to a HEG we would expect to get the best results in systems that are most like a HEG, such as a simple metal and the worst for very inhomogeneous systems like an isolated atom. In fact, the LDA approach has been surprisingly successful in a wide variety of systems. Part of the reason is that for most common material densities, the effects of exchange and correlation tend to be very short range and are well approximated by a local functional. Also, even though the shape of the exchange-correlation hole is qualitatively wrong, it still reproduces a good approximation to the spherical average which is the only factor that enters into the energy.

Nevertheless, the fact that we are using a local functional means that there cannot be an exact cancellation of the non-physical self-interaction of the Hartree energy operator - the so-called self-interaction error (SIE). In the context of the Sec. (2.4.3.4),

this spurious self-interaction is often associated with the lack of a derivative discontinuity in the functional. The physical consequence of this is to delocalize the electron wavefunctions which tend to artificially reduce bandgaps.

2.4.4.2 General gradient approximation

Mathematically, the next level of approximation would seem to naturally lead to considering not only the density, but the gradient of the density at a point as well:

$$E_{xc}^{GGA}[n] = \int n(r)\epsilon_{xc}(n(r), \nabla n)dr \tag{2.63}$$

Exchange and correlation are now semi-local functions, the degree of non-locality depending on the electron density. For high densities, exchange dominates correlation which is true in a physical system, but not in LDA. For a fixed density, as the gradient of the density increases, exchange interactions give a greater energy benefit (while correlation reduces). For most common densities, the non-locality of the exchange becomes so beneficial that GGA favors density inhomogeneity more than LDA. Only for low densities does correlation start to compare with exchange effects; as the density goes to zero the GGA exchange-correaltion hole eventually becomes as local as LDA.

GGA functionals tend to improve total energies and structural energy differences over LDA. They are more accurate in determining the binding energy of inhomogeneous systems like molecules which historically lead to the widespread acceptance of DFT in the chemistry community. They tend to favor non-uniform density more than LDA, which tends to soften bonds, an effect that sometimes corrects and sometimes over corrects the LDA prediction. However, as with LDA, the GGA functionals are still sensitive to the SIE that makes bandgap determination unreliable.

2.4.4.3 Hybrid functionals

The problem with local and semi-local functionals such as LDA and GGA are that they are too restrictive to reproduce all the effects of the 'true' exchange-correlation functional [9]; a local (or semi-local) exchange operator cannot completely cancel SIE. Hybrid functionals attempt to correct the self-interaction error inherent in LDA/GGA by mixing a portion of non-local exact HF exchange.

Hybrid functionals are constructed through the adiabatic connection approach given in Sec.(2.4.3.3). We assume a form of E_{xc} appropriate for independent particles and another for a fully coupled system and use a coupling constant λ to characterize the dependence of E_{xc}^{hyb} ($\lambda = 0$ completely independent particles, $\lambda = 1$ is fully coupled system). The independent particle exchange-correlation energy is just the HF exchange energy which is identical to the non-local operator in Eq.(2.46). For the fully interacting system, the exchange-correlation hole is deeper and thus more localized around the electron, so a local or semi-local functional should be accurate. So the fully coupled system is typically considered to be some form of local (LDA or GGA) functional. Based on how the $E_{xc}(\lambda)$ varies with λ Perdew, Ernzerhof, and Burke [10] proposes the following form:

$$E_{xc}^{hyb} = E_{xc}^{GGA} + \frac{1}{4}(E_x^{HF} - E_{xc}^{GGA}) \tag{2.64}$$

where 25% of the hybrid exchange-correlation energy comes from HF exchange and the rest from a local functional. This value of $\frac{1}{4}$ for the exchange mixing is semiempirical and based on minimizing the error of atomization energies for typical molecules (such as in the G1 data set) as well as the desireable property that the hybrid functional matches the GGA functional in value, slope, and second derivative at $\lambda = 1$

With this approach, there is still the problem that calculating the HF exchange is long-range and computationally expensive. Hinting at a solution is work of Kohn and others that shows that exchange interactions decay exponentially according to

bandgap in solids (and algebraically in metallic systems). To this end, a further refinement has recently been suggested by Heyd, Scuseria, and Ernserhof (HSE) where the exact exchange term is only calculated for short range interactions (where they are most critical) and we return to a purely local functional for longer ranges. To speed up the extinction of the HF exchange we introduce a screened potential (only for calculating the exchange interaction) based on splitting the Coulomb operator into short and long range components and only include HF exchange for the short range interactions:

$$\frac{1}{r} = \frac{\text{erfc(wr)}}{r} + \frac{\text{erf(wr)}}{r} \tag{2.65}$$

where w determines the range of the HF interations. For w=0 the second, long-range (LR) term becomes zero and the short-range (SR) term is exactly the full Coulomb potential and HF exchange is calculated everywhere. For very large values for w, the HSE functional tends toward the purely local GGA form (for HSE, we use the PBE-GGA functional).

The exchange-correlation energy has the following form:

$$E_{xc}^{HSE} = \alpha E_x^{HF,SR}(w) + (1-\alpha)E_x^{PBE,SR}(w) + E_x^{PBE,LR}(w) + E_c^{PBE} \tag{2.66}$$

where E_c and E_x are the individual correlation and exchange energies respectively. Thus the HSE functional can be viewed as a adiabatic connection functional for the short-range portion of the exchange, while the long-range exchange and correlation are treated as a semi-local functional.

HSE retains the very accurate energetics of previous hybrid functionals, and even improves upon the geometry optimization of the widely regarded most accrurate hybrid functional B3LYP. In addition, HSE is far less computationally expensive than previous hybrids in solid-state systems, especially for small-bandgap and metallic sys-

tems.

2.4.5 Periodic systems and planewave basis

A good choice of basis function is key to both an efficient representation of the electron orbitals in KS theory as well as accelerating the numerical calculations in the self-consistent loop of Fig.(2.1). In extended bulk systems (i.e. crystals) a convenient choice would take advantage of the periodicity of the system, since the wavefunctions must share the same periodicity as the lattice. A natural choice would be to use a Fourier series as the complete basis (i.e. planewaves). A periodic KS orbital may be written:

$$\phi_i(r) = \sum_q c_{i,q} \times \frac{1}{\Omega} e^{(iq\cdot r)} = \sum_q c_{i,q} \times |q\rangle \tag{2.67}$$

where $c_{i,q}$ are the coefficients of expansion of the KS orbital in the orthonormal plane wave basis and Ω is a normalization factor.

By substituting this into Eq.(2.24) we can express the KS SE in Fourier space as:

$$\sum_q \langle q'|\hat{H}_{eff}|q\rangle c_{i,q} = \epsilon_i c_{i,q'} \tag{2.68}$$

The effective potential in a crystal must also be periodic and can thus also be expressed as a Fourier series itself:

$$V_{eff}(r) = \sum_n V_{eff}(G_n) e^{iG_n\cdot r} \tag{2.69}$$

$$V_{eff}(G) = \frac{1}{\Omega} \int_\Omega V_{eff}(r) e^{-iG\cdot r} dr \tag{2.70}$$

with G_n are reciprocal lattice vectors. The matrix elements of the potential are:

$$\langle q'|V_{eff}|q\rangle = \sum_n V_{eff}(G_n)e^{i(q'+G_n)\cdot r}e^{iq\cdot r} \tag{2.71}$$

it is clear that the potential elements are non-zero only if q' and q differ by a reciprocal lattice vector G_n. If we define a vector k where $q = k + G_n$ and $q' = k + G_m$ then the KS SE for any k can be written as:

$$\sum_n H_{m,n}(k)c_{i,n} = \epsilon_i(k)c_{i,m}(k) \tag{2.72}$$

with matrix elements:

$$H_{m,n}(k) = \frac{\bar{h}^2}{2m_e}|k + G_m|^2\delta_{n,m} + V_{eff}(G_m - G_n) \tag{2.73}$$

For any given k, the solution is given by Eq.(2.67) summing only over $q = k + G$

$$\phi_{i,k}(r) = \frac{1}{\Omega}\sum_G c_{i,k}(G)e^{i(k+G)\cdot r} = u_{i,k}(r)e^{ik\cdot r} \tag{2.74}$$

which we recognize as Bloch's Theorem with $u_{i,k}(r)$ being the cell-periodic function.

Describing the KS single-particle orbitals with planewaves result in an especially simple form for the total energy [11]. Also, since real-space derivatives become multiplications in reciprocal space, we can use the fast Fourier Transform (FFT) to evaluate operators in the space they are diagonal, which simplifies evaluation of the kinetic and potential energy.

2.4.5.1 Infinities with periodic boundary conditions

The problem with periodic boundary conditions is that now the system is mathematically infinite in size which leads to divergences in the portions of the potential energy that are non-periodic such as the electron coulombic interactions (Hartree energy). Simply put, a single electron within a periodic system is interacting with an effectively

infinite distribution of charge, leading to an infinite energy.

The solution lies in the fact that we know that the system is overall neutral when we include the positive ions. Thus, we only need to consider the local charge density difference from the global average. Mathematically, if we consider the Fourier transform of the effective potential:

$$V_{eff}(G) = V_{ext}(G) + V_H(G) + V_{xc}(G) \tag{2.75}$$

this is equivalent to ignoring the $G = 0$ terms. For example, the Hartree energy can be related to the density via Poisson's equation:

$$-\nabla^2 v_H(r) = p(r) - \langle p \rangle \tag{2.76}$$

By taking the Fourier transform of both sides we arrive at:

$$V_H = \sum_{G \neq 0} \frac{n(G)^2}{G^2} \tag{2.77}$$

where the $G \neq 0$ term in the sum accounts for the subtraction of the average charge density. We then include the effects of the average density as an Ewald sum. The Ewald term is the energy of point ions in a compensating background and includes the effect of the ion-ion interactions as well as the average electron density with the ions and with itself.

2.5 Excited states and the GW method

In many instances, DFT is used to approximate excited state properties such as the bandgap and optical absorption spectra. However, as discussed previously, the KS single-particle wavefunctions and energies do not have physical meaning and makes no guarantee of any eigenvalue apart from the highest occupied orbital. A clue as to how to approach excited-state calculations comes from the way they are probed

in the lab. Electronic properties are usually investigated by measuring the response of a system to an external perturbation. Usually this involves either an electron (inverse photo-emission spectroscopy IPS) or a photon as a probe (photo-emission spectroscopy PES and optical absorption).

In PES photons of a specific energy are directed at a surface and electrons are liberated from the sample. By measuring the kinetic energy one can determine the properties of the occupied electronic states. In IPS the procedure is reversed; electrons with a certain energy are injected into a sample, filling unoccupied states and releasing photons which can then be collected to determine the characteristics of the unoccupied states. In both these procedures we are measuring single-body excitations of the system with the number of electrons N changing to $N+1$ or $N-1$. Thus, perturbation theory, using a Green's function which represents the addition and propagation of an electron in our system would seem a fruitful approach.

2.5.1 Second quantization

When dealing with interacting many-body systems it is useful to introduce a formalism known as second quantization. First quantization is the typical introduction to quantum mechanics where wavefunctions are expressed in terms of coordinates of particle 1, particle 2, and so on. However, for systems of identical particles it makes no sense to ask 'what is particle 1 doing?'. With first quantization we are forced to keep track of this unphysical information. Second quantization does not need to keep track of this information which makes it much more convenient for interacting many-body systems.

When dealing with independent particles we only need to consider wavefunctions that are antisymmetric under particle permutations. For products of single-particle states antisymmetrization can be handled very efficiently through use of Slater deter-

minants:

$$\Psi(x_1, x_2, ..., x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(x_1) & \phi_2(x_1) & ... & \phi_N(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & ... & \phi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(x_N) & \phi_2(x_N) & ... & \phi_N(x_N) \end{vmatrix} \tag{2.78}$$

where ϕ are single-particle wavefunctions. In fact, it can be shown that any N-electron system can be written as a linear combination of Slater determinants. In the form of Eq. (2.78) these determinants are expressed in a real-space basis. However, in quantum mechanics it is often useful to work in abstract states; instead of a real-space wavefunction $\psi_\alpha(x)$ we can talk in terms of an abstract state using Dirac notation $|\alpha\rangle$.

The Slater determinant for two electrons in states $\phi_\alpha(x)$ and $\phi_\beta(x)$ is just the antisymmetrized product of the two states:

$$\Psi(x_1, x_2) = \frac{1}{\sqrt{2}}(\phi_\alpha(x_1)\phi_\beta(x_2) - \phi_\alpha(x_2)\phi_\beta(x_1)) \tag{2.79}$$

We can also do this for Dirac states, a two-particle system specified by:

$$|\alpha, \beta\rangle = \frac{1}{\sqrt{2}}(|\alpha\rangle|\beta\rangle - |\beta\rangle|\alpha\rangle) \tag{2.80}$$

Within second quantization we specify these states using operators:

$$c_\beta^\dagger c_\alpha^\dagger |0\rangle = |\alpha, \beta\rangle \tag{2.81}$$

as long as these operators have certain qualities it can be shown that they can be used to represent Slater determinants.

In order to construct many-electron states we start from the no electron or vacuum state $|0\rangle$ which is assumed to be normalized. Then, for each single-electron state $|\alpha\rangle$ (again normalized) we introduce a *creation* operator $c_\alpha^\dagger$ which adds an electron in state

$|\alpha\rangle$ to an N-electron wavefunction, becoming an N+1 electron system. Applying this to the vacuum state yields:

$$c_\alpha^\dagger|0\rangle = |\alpha\rangle \tag{2.82}$$

In order to calculate the norm $\langle\alpha|\alpha\rangle$ we need to consider the adjoint of $c_\alpha^\dagger|0\rangle$ obtaining $\langle\alpha|\alpha\rangle = \langle 0|c_\alpha c_\alpha^\dagger|0\rangle$ which means the adjoint of the creation operator c_α must remove an electron from the state or else the state would no longer be normalized (i.e. there would be no overlap between $\langle 0|$ and $c_\alpha c_\alpha^\dagger|0\rangle$). Of course, you cannot remove an electron from the vacuum state, so $c_\alpha|0\rangle = 0$.

We further postulate that these operators change sign when exchanged:

$$c_\alpha^\dagger c_\beta^\dagger = -c_\beta^\dagger c_\alpha^\dagger \tag{2.83}$$

If these operators behave in the preceding way it can shown that they are equivalent to a Slater determinant of single-particle wavefunctions. We now have an intuitive picture of defining an arbitrary state using the vacuum state as a starting point, and then filling it up with particles in the desired single-particle states:

$$\Psi(x_1, x_2, ..., x_N) = |\alpha_1, \alpha_2, ..., \alpha_N\rangle = c_{\alpha_1}^\dagger c_{\alpha_2}^\dagger ... c_{\alpha_N}^\dagger|0\rangle \tag{2.84}$$

Here is one great benefit of second quantization - it provides a simple way to manage the antisymmetry of Slater determinants. The operators encode the proper sign for the antisymmetrized wavefunction in their relative positions.

To make a connection to the real-space picture there are a particularly useful type of operators called *field* operators $\hat{\Psi}^\dagger(x)$ which create an electron at a specific position and spin (there is also the corresponding annihilation field operator $\hat{\Psi}(x)$).

Any creation operator can be written in terms of the field operator:

$$c_\alpha^\dagger = \int dx \phi_\alpha(x) \hat{\Psi}^\dagger(x) \tag{2.85}$$

Furthermore, any multi-body operator can be written in terms of the field operators:

$$\hat{O} = \frac{1}{N!} \int dx_1 ... x_N \hat{\Psi}^\dagger(x_N) ... \hat{\Psi}^\dagger(x_1) O(x_1, ..., x_N) \hat{\Psi}(x_1) ... \hat{\Psi}(x_N) \tag{2.86}$$

Consequently, all operators are defined not just for a system with a specific N number of electrons, but for any number of electrons, which also is very useful when working with many-body systems.

2.5.2 Quasiparticle equation

Excitations in systems with strongly interacting particles can often be depicted in terms of weakly interacting *quasiparticles* [12,13]. A bare electron in a solid repels the other electrons in the system and can be thought of as being surrounded by a cloud of positive charge. The electron and this screening charge form an ensemble particle, or quasiparticle (QP), which only weakly interacts with the other quasiparticles through a screened Coulomb potential. The energy difference between the bare electron and quasiparticle is defined as the self-energy Σ_{XC} which accounts for all non-classical exchange and correlation effects in the system. Quasiparticle behavior is governed by the QP equation:

$$\left[-\frac{\nabla^2}{2} + V_{ext} + V_H \right] \Psi_{QP}(r) + \int dr' \Sigma_{XC}(r, r'; \epsilon_{QP}) \Psi_{QP}(r') = \epsilon_{QP} \Psi_{QP}(r) \tag{2.87}$$

where Ψ_{QP} and ϵ_{QP} are the QP wavefunctions and energies. In general, Σ_{XC} is non-local, non-hermitian, and energy dependent and cannot be determined exactly for

real systems.

Within many-body perturbation theory we can come up with an approach to calculate Σ_{xc} in terms of a single-particle Green's function. We start with the interacting hamiltonian in the second-quantization form:

$$\hat{H} = \int dr \hat{\psi}^{\dagger}(r,t) h_0(r) \hat{\psi}(r,t) + \frac{1}{2} \int dr dr' \frac{1}{|r-r'|} \hat{\psi}^{\dagger}(r,t) \hat{\psi}^{\dagger}(r',t) \hat{\psi}(r',t) \hat{\psi}(r,t) \quad (2.88)$$

where $\hat{\psi}^{\dagger}(r,t)$ and $\hat{\psi}(r,t)$ are the creation and annihilation field operators, $h_0 = \frac{1}{2}\nabla^2 + V_{ext}(r)$ is the one-body part of the hamiltonian. The ground state of the N-electron system is denoted as $|N\rangle$. We can define the one-body Green's function $G(r_1 t_1, r_2 t_2)$ as:

$$G(r_1 t_1, r_2 t_2) = -i\langle N|T[\hat{\psi}(r_1,t_1)\hat{\psi}^{\dagger}(r_2,t_2)]|N\rangle \quad (2.89)$$

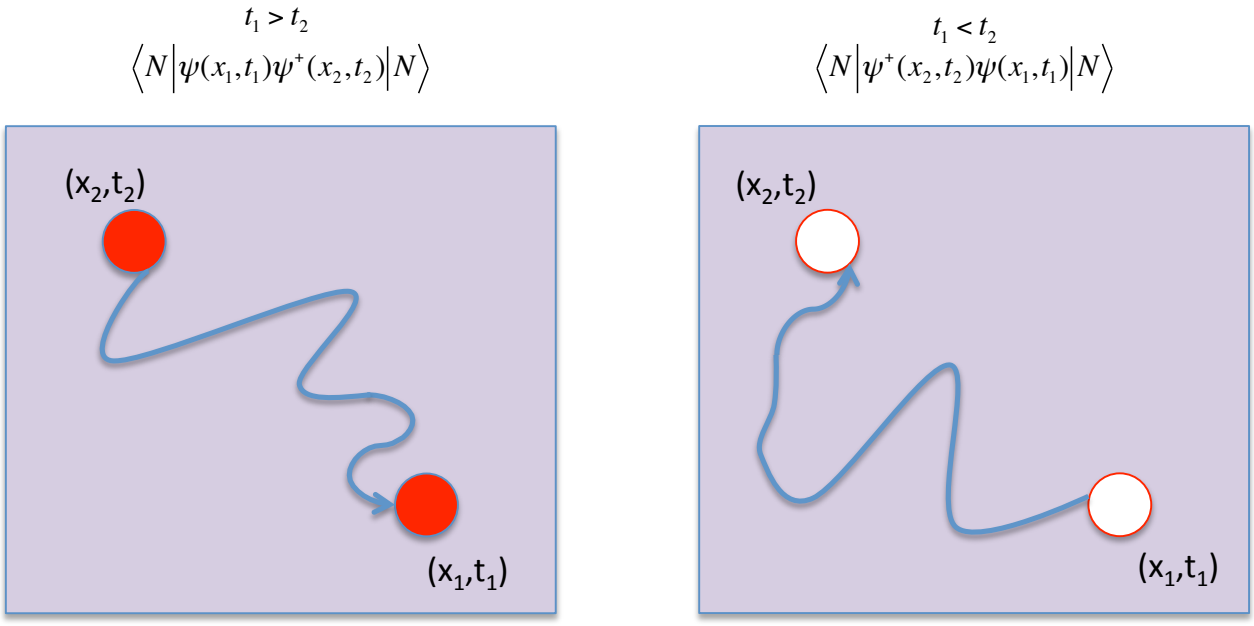
where T is the time-ordering operator. The physical interpretation of the one-particle Green's function is the probability amplitude to detect an electron at point r_1 and time t_1 given that an electron has been added to the system at point r_2 and time t_2 (if $t_1 > t_2$), otherwise this is the probability of detecting a *hole* at r_1 and time t_1) - see Figure(2.2).

Next, using the Heisenberg equation of motion (EOM) for the field operator:

$$i\frac{d\hat{\psi}(rt)}{dt} = [\hat{\psi}(r,t), \hat{H}] \quad (2.90)$$

we can derive the EOM for the Green's function as:

Figure 2.2: Single particle Green's function represents the probability to detect an electron at point r_1 and time t_1 given an electron has been added at point r_2 and time t_2 (if $t_1 > t_2$ as in left diagram), otherwise it represents a similar propagation of a electron hole (right diagram).



$$\left[i\frac{d}{dt_1} - h_0(r_1)\right] G(r_1t_1, r_2t_2) - i\int dx_3dt_3v(r_1, r_3)\langle N|T[\hat{\psi}^\dagger(r_3, t_1)\hat{\psi}(r_3, t_1)\hat{\psi}(r_1, t_1)\hat{\psi}^\dagger(r_2, t_2)]|N\rangle$$
$$= \delta(r_1 - r_2)\delta(t_1 - t_2) \tag{2.91}$$

where $v(r_1, r_3) = \frac{1}{|r_1-r_3|}\delta(t_1 - t_3)$ is the bare Coulomb interaction. If we recognize the integral contains the two-body Green's function:

$$G_2(r_1t_1r_2t_2; r_3t_3r_4t_4) = -\langle N|T[\hat{\psi}(r_1, t_1)\hat{\psi}(r_2, t_2)\hat{\psi}^\dagger(r_3, t_3)\hat{\psi}^\dagger(r_4, t_4)]|N\rangle \tag{2.92}$$

we see that the EOM for the one-body Green's function involves the two-body Green's function:

$$[i\frac{d}{dt_1} - h_0(r_1)]G(1, 2) + i\int d3v(1^+, 3)G_2(1, 3; 2, 3^+) = \delta(1, 2) \tag{2.93}$$

where we have adopted Hedin's simplified notation: $1 \equiv (r_1, t_1)$ and $1^+ \equiv (r_1, t_1 + \eta)$ where η is a positive infinitesimal introduced to ensure proper time-ordering of the field operators. A similar EOM can be dervied for G_2 depending on the three-body Green's function; this continues, leading to a hierarchy of equations.

We can separate the EOM into non-interacting (first term in Eq. 2.93) and the interacting (integral term) parts. We assume that the non-interacting term can be solved exactly, thus defining the independent-particle Green function G_0 for the non-interacting hamiltonian h_0:

$$[i\frac{d}{dt_1} - h_0(r_1)]G_0(1, 2) = \delta(1, 2) \tag{2.94}$$

Now considering the two-body Green's function, we do not know how the two particles propagate, but a good guess is to allow each particle to propagate independently

according to one-body Green's functions:

$$G_2(1,3;2,3^+)\delta(t_1^+ - t_3) = [G(1,2)G(3,3^+) + G(1,3^+)G(3,2)]\delta(t_1^+ - t_3) \tag{2.95}$$

The first term is the direct and the second is the exchange term. If we use only the direct term to approximate G_2 the EOM (2.93) becomes:

$$\left\{[\frac{d}{dt_1} - h_0(r_1)] + i\int d3v(1^+,3)G(3,3^+)\right\}G(1,2) = \delta(1,2) \tag{2.96}$$

which is an independent particle like equation with an added potential which is equivalent to the Hartree potential:

$$\begin{aligned} -i\int d3v(1^+,3)G(3,3^+) &= \int dr_3 \frac{n(r_3,t_1)}{r_1 - r_3} = V_H(1) \\ [i\frac{d}{dt_1} - h_0(r_1) - V_H(1)]G(1,2) &= \delta(1,2) \end{aligned} \tag{2.97}$$

We can then take both the direct and exchange terms as the approximation to G_2 resulting in an EOM:

$$[\frac{d}{dt_1} - h_0(r_1) - V_H(1)]G(1,2) + i\int d3v(1^+,3)G(1,3^+)G(3,2) = \delta(1,2) \tag{2.98}$$

where the interaction term is now a non-local operator corresponding to the exchange interaction in the Hartree-Fock approximation. We could continue on like this with the three particle Green's function and so on, but this rapidly becomes unmanageable. It is much more convenient to introduce the concept of the self-energy. We assume that we have solved the infinite series of EOM's and look for a solution in the form:

$$[\frac{d}{dt_1} - h_0(r_1) - V(1)]G(1,2) + i\int d3\Sigma(1,3)G(3,2) = \delta(1,2) \tag{2.99}$$

with Σ the self-energy operator which includes all interaction effects and $V(1) =$

$\phi(1) + V_H(1)$ with $\phi(1)$ being an external potential which acts as a probe of the system. A physical intuition of the self-energy is more obvious if we transform the EOM into the frequency or energy domain:

$$[w - h_0(r_1) - V(r_1, w)]G(r_1, r_2, w) - \int dr_3 \Sigma(r_1, r_3; w)G(r_3, r_2; w) = \delta(1, 2) \quad (2.100)$$

in matrix notation:

$$\begin{aligned} (w1 - h_0 - V)G - \Sigma G = 1 \\ G^{-1} = w1 - h_0 - V - \Sigma \end{aligned} \quad (2.101)$$

if we compare this to the non-interacting EOM (without Σ):

$$G_0^{-1} = w1 - h_0 - V \quad (2.102)$$

we can write:

$$G^{-1} = G_0^{-1} - \Sigma \quad (2.103)$$

The poles of G are shifted in energy by Σ compared to G_0. This equation can also be written:

$$G = G_0 + G_0 \Sigma G \quad (2.104)$$

which is known as the Dyson equation.

To this point we have defined everything in terms of the single-body Green's function. The Green's function itself can be defined in terms of the quasi-particle wavefunctions and energies. As a direct derivation from the Green's function theory of differential equations, the solution $G(r_1, r_2, w)$ for Eq.(2.100) can be found as:

$$G(r_1, r_2, w) = \Sigma_n \frac{\Psi_{QP}(r_1, w)\Psi^*_{QP}(r_2, w)}{w - \epsilon_{QP}(w)} \tag{2.105}$$

where Ψ_{QP} and ϵ_{QP} are the QP wavefunctions and energies from Eq.(2.87).

2.5.3 Hedin's equations

Hedin's equations lead to a perturbation expansion of the self-energy in terms of the screened interaction W rather than the bare Coulomb interaction [12, 14]. This is critical in obtaining a converging series for the self-energy. Central to Hedin's approach is realizing that the two-body Green's function appearing in the self-energy can be related to the response of the one-body Green's function with respect to a small external potential $\phi(rt)$ introduced into Eq.(2.93) adiabatically:

$$\frac{dG(1,2)}{d\phi(3)} = G(1,2)G(3,3^+) - G_2(1,2,3,3^+) \tag{2.106}$$

substituting into Eq.(2.99) we arrive at:

$$\Sigma(1,2) = i \int dr3dr4v(1,3)\frac{dG(1,4)}{\phi(3)}G^{-1}(4,2) \tag{2.107}$$

We next define the inverse dielectric function ϵ^{-1} at the change in potential V due to a small variation in the external potential ϕ:

$$\epsilon^{-1}(1,2) = \frac{dV(1)}{d\phi(2)} \tag{2.108}$$

and we can now write:

$$\begin{aligned} V(1) &= \phi(1) + \int dr_3 v(1,3)n(3) \\ dV(1) &= d\phi(1) + \int dr_3 v(1,3)dn(3) \\ \epsilon^{-1} &= d(1,2) + \int dr_3 v(1,3)\frac{dn(3)}{d\phi(2)} \end{aligned} \tag{2.109}$$

where we can define the polarizability as the change in density upon a change in the total potential $P(3,2) = \frac{dn(3)}{dV(2)}$. We can define the screened Coulomb potential W:

$$W(1,2) = \int dr_3 \epsilon^{-1}(1,3) v(3,2) \tag{2.110}$$

The polarizability can be expanded to:

$$P(1,2) = \frac{dn(2)}{dv(2)} = -i\frac{dG(1,1^+)}{dV(2)} = i\int dr_3 dr_4 G(2,3) - \Gamma(3,4,1)G(4,2^+) \tag{2.111}$$

where $\Gamma = -\frac{dG^{-1}(3,4)}{dV(1)}$ is variation in the self energy with respect to the total potential which is termed the vertex function. This can also be expressed in terms of the non-interacting Green's function and the self energy:

$$\Gamma(1,2,3) = -\frac{dG^{-1}(1,2)}{dV(3)} = -\frac{dG_0^{-1}(1,2)}{dV(3)} + \frac{\Sigma(1,2)}{dV(3)} \tag{2.112}$$

which is equivalent to:

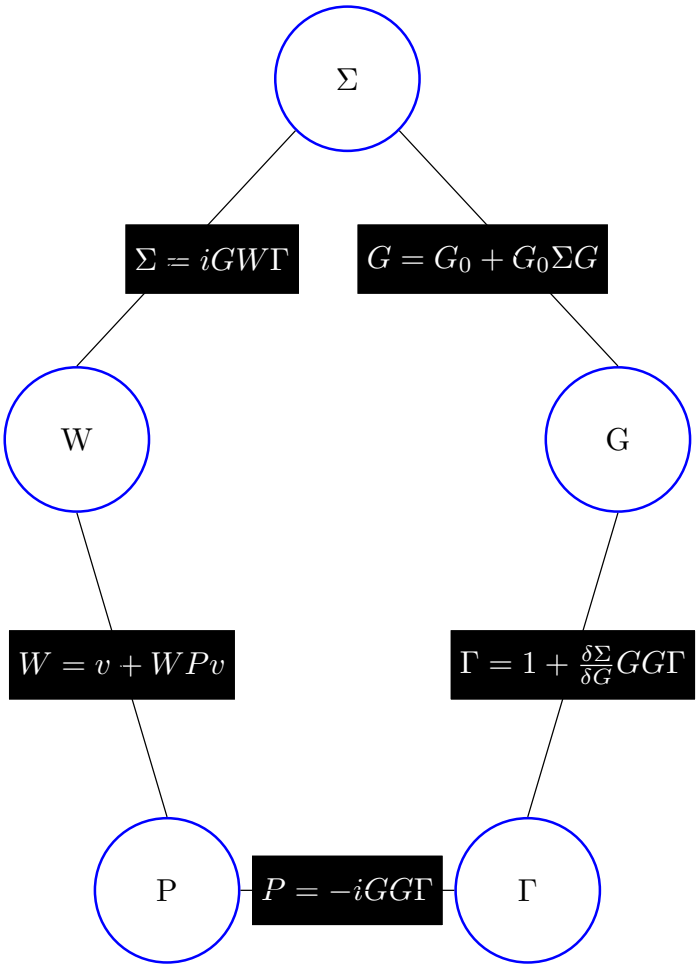
$$\Gamma(1,2,3) = \delta(1,2)\delta(1,3) + \int dr_4 dr_5 dr_6 dr_7 \frac{d\Sigma(1,2)}{dG(4,5)} G(4,6)G(7,5)\Gamma(6,7,3) \tag{2.113}$$

After further manipulation, we arrive at an expression for the self-energy in terms of the screened potential W:

$$\Sigma(1,2) = -\int dr_3 dr_4 G(1,4) W(1^+,3)\Gamma(4,2,3) \tag{2.114}$$

Eq.(2.110 2.111 2.113 2.114) together are Hedin's equations, which along with Dyson's equation of Eq.(2.104) form a closed set of coupled equations - see Figure(2.3). Sadly, these equations are not solvable for even the simplest systems such as the HEG. Their real utility is as the starting point for a many-body perturbation method based on the screened Coulomb potential W.

Figure 2.3: Schematic representation of Hedin's equations. These are a set of coupled integral equations which relate the self energy Σ, Green's function G, vertex function Γ, polarizability P, and screened Coulomb interaction W.



2.5.4 GW approximation

One potential strategy for solving Hedin's equations would be to start by setting the self-energy to zero [15]. The one-particle Green's function simplifies to G_0 and the other functions reduces to:

$$\begin{aligned} \Gamma(1,2,3) &= \delta(1,2)\delta(1,3) \\ P(1,2) &= -iG(1,2)G(2,1) \\ \Sigma(1,2) &= iG(1,2)W(1,2^+) \end{aligned} \tag{2.115}$$

the expression for the self-energy is iGW, hence the name of the approximation. G_0 is provided by Eq.(2.105) using the eigenfunctions and eigen-energies of some reference single-particle hamiltonian. Since the GW method is based on perturbation theory, the reference hamiltonian should be chosen to give the best approximation for G and W. The reference hamiltonian is often provided by KS-DFT:

$$G_0(r_1, r_2, w) = \Sigma_{nk} \frac{\phi_{nk}^{KS}(r_1)\phi_{nk}^{KS}(r_2)^*}{w - \epsilon_{nk}^{KS} + i\eta sgn(\epsilon_{nk}^{KS} - \mu)} \tag{2.116}$$

Though there is no direct relation between the KS single-particle energies and QP excitations, the KS values often approximate the QP band structure well and excellent results have used them as the starting point for GW calculations.

The polarizability is taken within the independent-particle or random phase approximation (RPA):

$$P_0(r_1, r_2, w) = -\frac{i}{2\pi} \int G_0(r_1, r_2, w - w')G_0(r_2, r_1, w)e^{-i\eta w'} dw \tag{2.117}$$

$$\epsilon(r_1, r_2, w) = \delta(r_1, r_2) - \int dr_3 \frac{P_0(r_1, r_3, w)}{|r_3 - r_2|} \tag{2.118}$$

The screened Coulomb interaction is given by:

$$W_0(r_1, r_2, w) = \int dr_3 \frac{\epsilon^{-1}(r_3, r_2, w)}{|r_1 - r_3|} \tag{2.119}$$

and the self-energy:

$$\Sigma(r_1, r_2, w) = \frac{i}{2\pi} \int G_0(r_1, r_2, w - w') W_0(r_1, r_2, w) e^{-i\eta w'} dw' \tag{2.120}$$

In principle, this process can be repeated until self-consistency is achieved. However, this is unwieldy in practice and one often stops after the first round - the G_0W_0 approximation.

Another approach would be to solve the quasi-particle equation directly rather than for the one-body Green's function. There is a similarity between the quasi-particle equation Eq.(2.87) and the KS DFT eigenvalue problem:

$$\left[-\frac{\nabla^2}{2} + V_{ext} + V_H \right] \Psi_{KS}(r) + V_{XC} \Psi_{KS}(r) = \epsilon_{KS} \Psi_{KS}(r) \tag{2.121}$$

This suggests that we could treat the difference between Σ_{XC} and V_{xc} as a perturbation on top of the KS calculations. In fact, the approximation $\Psi_{QP}(r) \simeq \Psi_{KS}(r)$ is reasonable in many instances so the quasi-particle energy can be expressed as:

$$\epsilon_{QP} = \epsilon_{KS} + \langle \phi_{KS} | \Sigma(r, r', w = \epsilon_{QP}) - V_{XC}(r) | \phi_{KS} \rangle \tag{2.122}$$

As Σ depends on the energy, this should be solved self-consistently. In practice Σ is often linearised to simplify calculations.

In summary, with either approach to the GW calculations, DFT is usually a good choice as a starting point. In this way the initial approximation to the Green's function can be derived directly from DFT results via Eq.(2.116). Then we can use Hedin's equations to self-consistently solve for the self energy. Being a perturbative

method, the GW technique, though it will eventually converge to the correct results, in practice does depend on our initial guess for the system configuration. We have found that different exchange-correlation functionals will affect results such as fundamental bandgap to some degree. However, even a single self-consistent loop (G_0W_0) using LDA functionals results in very accurate bandgaps for many systems.

2.6 Summary

In this chapter we have reviewed the general problem of solving the Schrödinger Equation for complex physical systems. We have chosen to work within the Density Functional Theory framework which recasts the problem into a simpler picture of interacting densities rather than multi-body wavefunctions. The simplifications made and their impact on the quality of our solution have been highlighted in order to set up the following chapters, which deal intimately with overcoming these limitations in real-world applications.

Chapter 3

Ab Initio study of intrinsic defects and dopability of zinc phosphide

Zinc Phosphide (Zn_3P_2) could be the basis for cheap and highly efficient solar cells. Its use in this regard is limited by the difficulty in n-type doping the material. In an effort to understand the mechanism behind this, the energetics and electronic structure of intrinsic point defects in zinc phosphide are studied using generalized Kohn-Sham theory and utilizing the Heyd, Scuseria, and Ernzerhof (HSE) hybrid functional for exchange and correlation. Novel 'perturbation extrapolation' is utilized to extend the use of the computationally expensive HSE functional to this large-scale defect system. According to calculations, the formation energy of charged phosphorus interstitial defects are very low in n-type Zn_3P_2 and act as 'electron sinks', nullifying the desired doping and lowering the fermi-level back towards the p-type regime. This is consistent with experimental observations of both the tendency of conductivity to rise with phosphorus partial pressure, and with current partial successes in n-type doping in very zinc-rich growth conditions.

3.1 Introduction

Zinc Phosphide has great potential as a photovoltaic material; it absorbs strongly in the visible spectrum ($> 10^4 - 10^5 cm^{-1}$), has long minority carrier diffusion lengths (5-10 μm), and a direct, almost ideal bandgap ($\sim$1.5eV) [16]. Both zinc and phosphorus are Earth-abundant elements, greatly aiding their widespread use. However, no

practical means of creating Zn_3P_2 crystals with n-type doping has been found. This has prevented the typical p-n homojunction solar cells and current implementations instead rely upon metal-semiconductor junctions or p-n semiconductor heterojunctions [17]. The best results are currently with p-Zn_3P_2/Mg Schottky diode where the maximal open-circuit voltage ($\sim$0.5eV) is the limiting factor on the efficiency of these devices (currently 6%) [18].

Differences in growth environment has been shown to affect the electrical properties significantly [19]. Specifically, resistivity ranging over 100 to $10^5\Omega$-cm have been measured in single and polycrystalline samples. Such a large range of values may point to an intrinsic defect mechanism dominating carrier concentrations. Despite many years of research, the exact nature of the various intrinsic defects as well as their small-scale structure and properties are only partially known. So, as a step towards understanding the n-type doping difficulties, it is natural to first explore the role of intrinsic defects in this system.

Density Functional Theory (DFT) is the method of choice for studying the electronic structure of defects in semiconductors. It is the only *ab initio* approach currently tractable for calculation of the energetics of cells of sufficient size to explore isolated defects (at least on the order of 100's of atoms). For computational simplicity, the most common exchange-correlation functionals utilized are the local-density approximation (LDA) and generalized-gradient approximation (GGA). However, while the defect-cell total energies are typically well-described by LDA or GGA, the well-known bandgap problem of these methods poses problems in determining the precise electronic structure of the defects themselves. Subsequently, this can lead to different predictions of defect stability amongst calculations even on identical systems. For example, recent work of various groups on ZnO has lead to predictions that the oxygen vacancy acts as both a shallow and deep defect [20, 21]. Recently, more accurate techniques such as hybrid functionals [22] and GW excited-state calculations [12] yield greatly improved bandgap prediction. However, these techniques are still too

expensive for the large scales required of defect systems.

Regardless of the exact *ab initio* procedure chosen, there is a fundamental problem in studying defects with a necessarily small amount of atoms (even the most efficient approach of LDA-based DFT is limited to on the order of hundreds of atoms); that is, we are imposing a degenerate doping condition, which is seldom the regime of interest (typical defect or dopant concentrations rarely exceed parts-per-million). As such, care must be taken to correct for interactions that are artifacts of this incorrectly high defect concentration such as image charges and spurious hybridization. In this chapter we detail the methods that allow us to compute defect levels and energetics separate from these effects.

3.2 Thermodynamics of Defect Stability

Of central concern is the stability of the intrinsic defects in Zn_3P_2 and their effect on the fermi level. The stability of a defect is largely determined by its formation energy; in the supercell formalism put forth by Lany and Zunger [23] the formation energy of a defect with charge q is given by a sum of three terms:

$$\Delta H_f = [E_D - E_H] + q(E_V + \Delta E_f) + \sum_\alpha n_\alpha(\mu_\alpha^o + \Delta\mu_\alpha) \tag{3.1}$$

E_H and E_D are the total energy of the perfect host and host+defect supercells respectively. The first term on the right is the difference in bond energy brought about by the defect.

The second term in (3.1) represents the energy cost of exchanging electrons with the 'electron reservoir'. E_V is the reference energy of the reservoir and is the price we pay for removing an electron from the top of the valence band (ie. the energy of a hole at the Valence Band Maximum) [24]. Consequently, equation (3.1) describes

the energetics of forming a defect while conserving charge [23]. The calculated total energy (E_H) of the system follows Janak's Theorem [25]:

$$\frac{dE_H(n_i)}{dn_i} = e_i \tag{3.2}$$

where n_i is the occupation of the highest occupied state i with eigenvalue e_i. For an infinite system e_i is identical to E_V. Thus, E_V can be calculated as the energy difference between a host and a host+hole cell in the limit that the number of electrons (N) tends to infinity:

$$\lim_{N\to\infty}[E_H(N) - E_H(N-1)] = E_V \tag{3.3}$$

As a practical matter, a good approximation can be attained for relatively small systems - in the present work the difference between a neutral perfect supercell and a supercell and a hole ($E_H(0)$ - $E_H(+)$) is used. We now have a good approximation for the electron chemical potential at the VBM. Finally, ΔE_f is the additional energy of electrons in our system above the VBM and is the proxy for specifying the doping regime (p or n-type) of the bulk.

The crystal growth environment affects the formation energy via the chemical potentials ($\mu_\alpha = \mu_\alpha^o + \Delta\mu_\alpha$) in equation (3.1). These represent the energy cost of exchanging atoms with the chemical reservoir. By convention, the formation energies are defined in relation to the standard states of the constituents of a system. Thus, the elemental chemical potentials are broken into two parts: μ_α^o is the chemical potential of the standard state of the element and $\Delta\mu_\alpha$ is the chemical potential of the element in relation to the standard state. The growth conditions are reflected in $\Delta\mu_\alpha$ (ie. a maximally rich growth environment of a certain element would have $\Delta\mu_\alpha = 0$), it becomes more negative for lower concentrations of an element during crystal formation. The chemical potentials are added or subtracted from the formation energy according to the number of atoms of a certain species is deposited or withdrawn from the growth reservoir (n_α); it is +1 if an atom is added (ie. vacancy defects), if we remove an atom it is -1 (ie. interstitial defects).

Thus, the formation energy is determined as a function of fermi-level (ΔE_f) for a specific growth condition (dictated through $\Delta\mu_\alpha$). Assuming equilibrium growth conditions, the chemical potentials are restricted to values that maintain a stable compound and don't permit competing phases to exist. For Zn_3P_2 we have the following constraints:

In order not to precipitate the elemental form of Zn or P, we must have:

$$\mu_{Zn} \leq \mu^o_{Zn};\ \mu_P \leq \mu^o_P \tag{3.4}$$

or, equivalently:

$$\Delta\mu_{Zn} \leq 0;\ \Delta\mu_P \leq 0 \tag{3.5}$$

where $\Delta\mu_{Zn}$=0 would define the maximumly-rich Zn growth condition.

The stability condition for Zn_3P_2 is:

$$3\Delta\mu_{Zn} + 2\Delta\mu_P \leq \Delta H_f(Zn_3P_2) \tag{3.6}$$

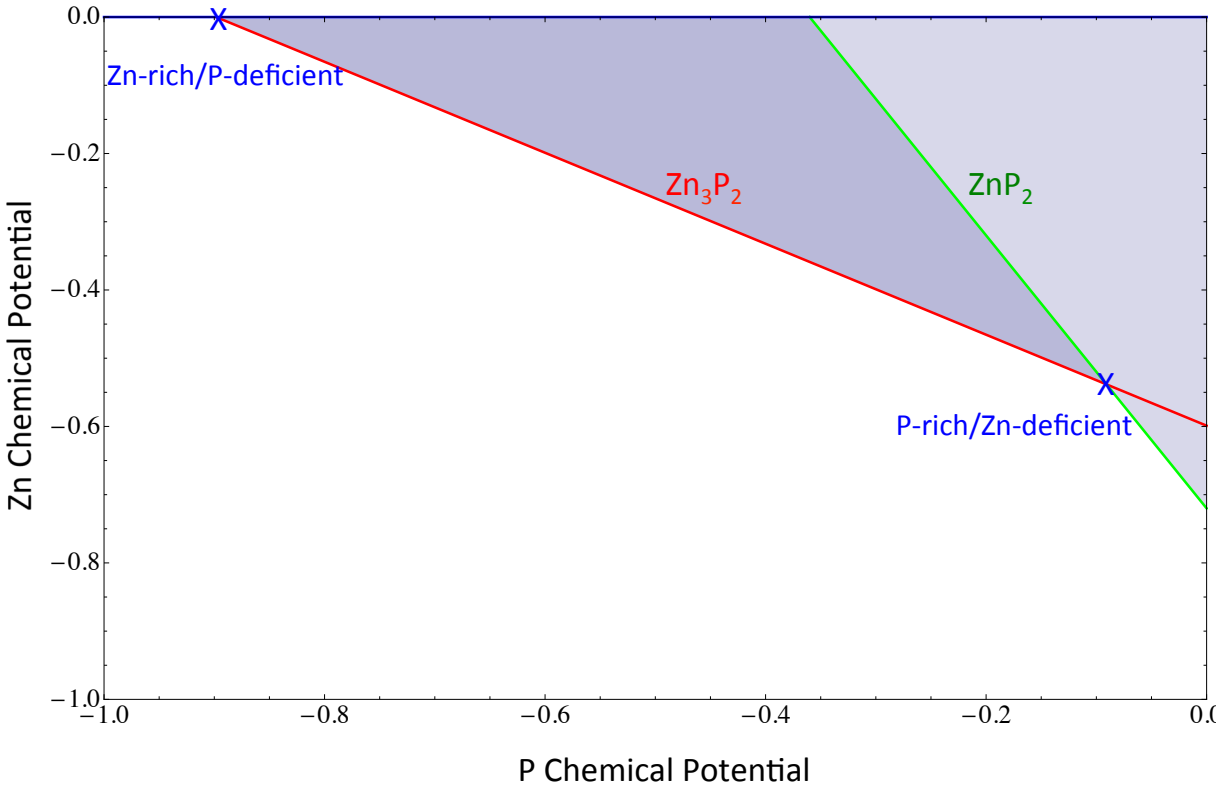
The only other competing phase is ZnP_2 (phosphorus-rich phase) and avoiding its formation yields the following restriction:

$$\Delta\mu_{Zn} + 2\Delta\mu_P \geq \Delta H_f(ZnP_2) \tag{3.7}$$

There is a further issue of incomplete error cancellation in DFT when energy differences are taken between chemically dissimilar systems [26] (such as between compounds and their elemental constituents). Here we correct the chemical potentials according to [26] which results in better agreement between predicted and experimental formation energies (1.04 eV non-corrected, 1.79 eV corrected, 1.5 eV experimental).

Combining all the conditions described by Eq 5-7, we can determine an allowed

Figure 3.1: Zn_3P_2 allowed chemical potentials. The red and green lines are the stability limits for Zn_3P_2 and ZnP_2 respectively. The dark purple region corresponds to the range of chemical potentials where we are assured of forming zinc phosphide crystals.



region of chemical potentials where we are assured of forming only Zn_3P_2 - see Figure 3.1. Later, when we discuss formation energies for a specific growth environment it will be this figure which defines the chemical potentials to use for the Zn-rich/P-poor versus the Zn-poor/P-rich regimes.

3.3 Non-Degenerate doping corrections to supercell formalism

Although the use of a supercell geometry within a DFT framework is a common approach for defect calculations, care must be taken to avoid spurious finite-size effects as well as known deficiencies in DFT's ability to model excited state energies. There are numerous examples of conflicting theoretical predictions due to incomplete or inconsistent corrections for these effects (as in the previously mentioned ZnO work). Here we will discuss the major aspects of the corrections.

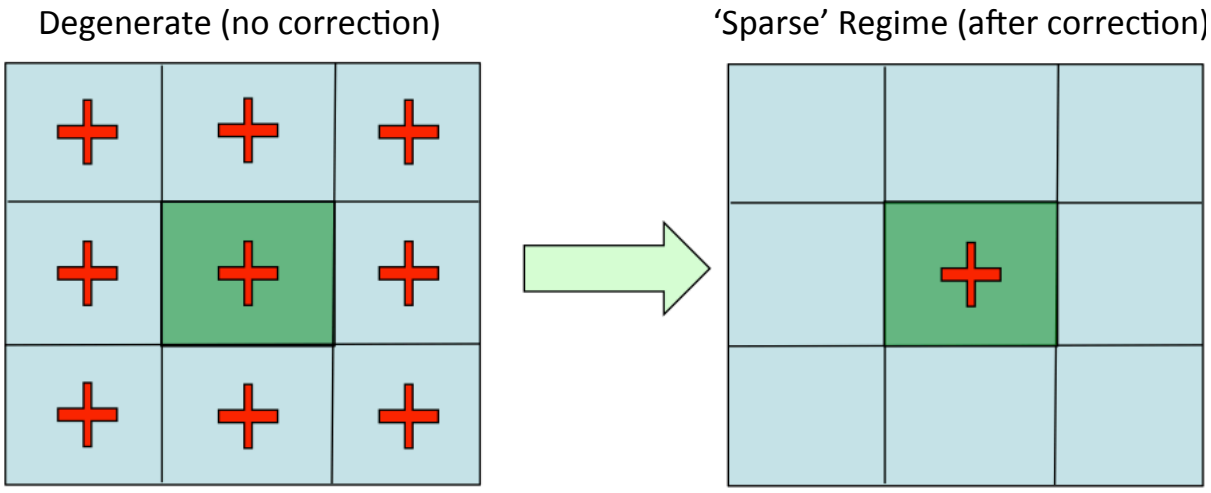
3.3.1 Finite Size Effects

With current limitations on computing power, DFT calculations are typically restricted to cells on the order of hundreds of atoms. Even the addition of a single point defect would thereby result in defect concentrations on the order of tenths of a percent, normally describing degenerate conditions. Usual semiconductor defect concentrations are much more dilute (on the order of parts-per-million), and often result in far different material properties than within the degenerate regime. Due to the low dielectric strength (and hence poor screening) of Zn_3P_2 these effects are especially troublesome in this study. However, if we are careful about correcting our results, we can still make accurate predictions about what the dilute environment should look like. The three main finite-size corrections are enumerated below.

3.3.1.1 Image Charge Correction

The drawback of the use of a standard supercell geometry is that defects are periodically and infinitely repeated spatially. The defect, instead of being surrounded by a large region of perfect bulk crystal as it would be under non-degenerate conditions, is now surrounded by mirror images of itself (Fig. 3.2). This will result in somewhat

Figure 3.2: Image charge correction schematic. The supercell (in dark green) is periodically replicated in space, due to planewave basis, leading to artificially high defect concentrations. For charged defects this results in an overestimation of the electrostatic energy in the system. We want to correct for this to recover the 'sparse' level of defects.



frustrated ionic relaxation, though these elastic energy effects tend to be short range and is rarely a problem for even modestly sized cells (there is very little difference in relaxation energies for even a 2x2x2 supercell versus a unit cell of Zn_3P_2). However, when dealing with charged defects we form 'image charges' leading to spurious electrostatic interactions. These coulombic interactions between the defect and its mirror charges are long-ranged and significant even for large cells.

Corrections for this 'image charge' effect have been the subject of much research, though the most common approach is based on the work of Makov and Payne [27]. They considered the charge density to be the contribution of the periodic charge of

the underlying crystal structure and the charge density of the aperiodic defect (which is simply the electron density difference between the host and host+defect cells). The multipole correction to the formation energy is:

$$E_{IC} = \frac{q^2 \alpha_M}{2\epsilon L} + \frac{2\pi q Q_r}{3\epsilon L^3} + O(L^{-5}) \tag{3.8}$$

where α_M is the supercell lattice-dependant Madelung constant, L is the length of the supercell, ϵ is the static dielectric constant, and Q_r is the second radial moment of the aperiodic charge density. The first two terms are the monopole and quadrupole corrections respectively; the quadrupole correction typically $\sim 30\%$ of the monopole term [25].

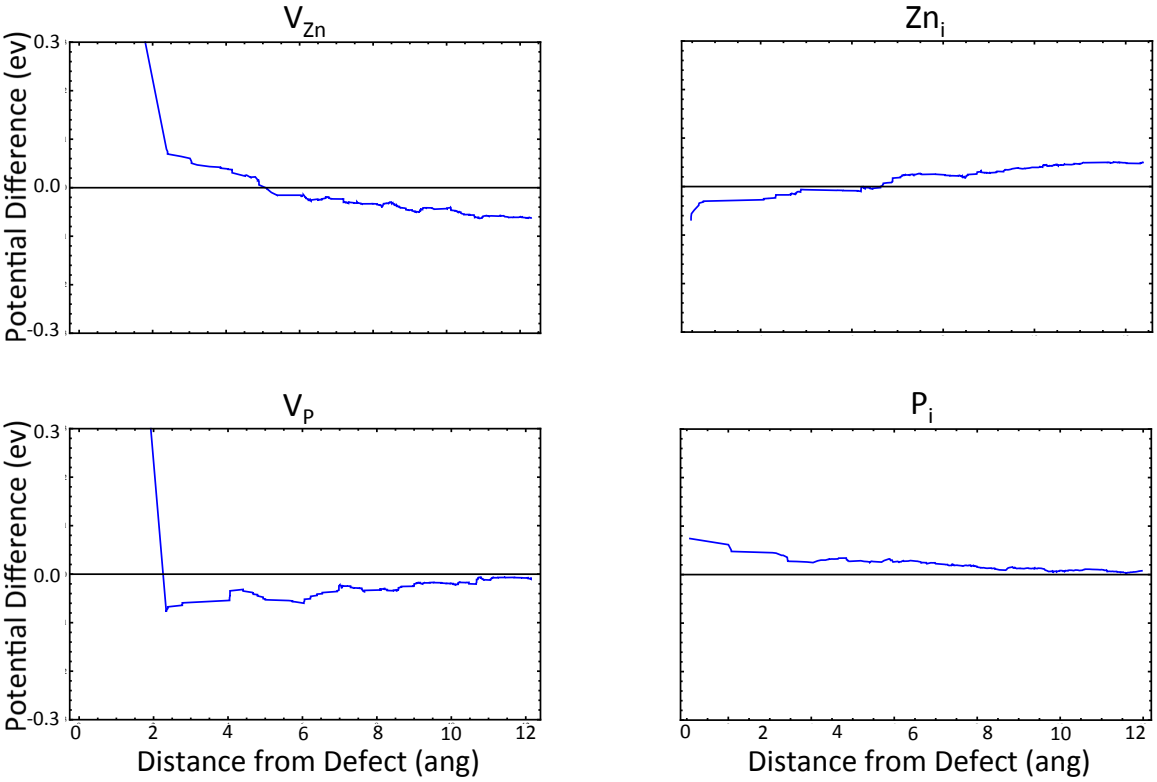
3.3.1.2 Potential Alignment Correction

In the case of charged defects with periodic boundary conditions there is a violation of charge neutrality, which causes the Coulomb potential to diverge [28]. In momentum-space formalism, one usually sets the G=0 term of the electrostatic and ionic potential (V_H(G=0) and V_I(G=0)) to zero. The Kohn-Sham eigenvalues are thus only defined with respect to the average electrostatic potential of the cell. For neutral systems this arbitrary offset still leads to a well-defined total energy since the electron-electron and ion-ion contributions exactly cancel. In a charged system, ignoring the G=0 term can be viewed as equivalent to a uniform background charge (jellium) compensating for the net charge - though it is important to note that this only occurs for the potential. In a charged cell there is now an arbitrary offset to the total energy. The charged cell energies E_D and $E_H(+)$ in Eq. (3.1) have to be compensated for in order to treat them on an equal footing with the neutral cell [29]. The potential-alignment correction is:

$$E_{PA} = q(V^r_{D,q} - V^r_H) \tag{3.9}$$

where the charged defect ($V^r_{D,q}$) and host reference (V^r_H) potentials are atomic sphere-averaged electrostatic potentials far from the defect site, q is the charge of the defect-

Figure 3.3: Potential alignment correction. Differences in electrostatic potential between perfect and defect cells far from defect area are used to correct charged cell energies.



see Figure 3.3.

3.3.1.3 Band-filling Correction

The artificially degenerate doping regime can cause defects to form bands rather than isolated states within the bandgap [30, 31]. For shallow defect states, this incorrect dispersion can cause abnormally large hybridization with the extended band states (either conduction or valence) and subsequently partially populate these bands. In order to obtain accurate defect cell energies (E_D) we have to correct for the extra energy in the system due to electron populations at these higher (or lower) energies. For shallow donors, the correction is:

$$E_{BF} = -\sum_{n,k} \Theta(e_{n,k} - e_C)(w_k f_{n,k} e_{n,k} - e_C) \tag{3.10}$$

Where $e_{n,k}$ is the k-dependent energy of state at band index n, e_C is the conduction band minimum energy of the defect-free bulk after potential alignment, $f_{n,k}$ is the band occupation, and w_k is the k-point weight. Θ is the Heaviside step function.

3.3.2 Band-gap Error

The most common exchange-correlation potentials, LDA and GGA, severely underestimates the bandgap of most semiconductors. This has two damaging effects on defect calculations. The defect-induced states may lie artificially close in energy to some band states causing excessive hybridization and ambiguity between shallow or deep defect behaviour. Secondly, the range of electron chemical potentials used to calculate defect formation energies will be too small, possibly incorrectly predicting unstable charged defect states.

Fundamentally, the bandgap error in LDA or GGA is the result of a lack of continuity with respect to the number of electrons in the exchange-correlation potential [7]. This in turn leads to self-interaction error (SIE) associated with a bias towards delocalized wavefunctions. The most common means to correct this situation is to add a Hubbard-like potential to penalize partial state occupancies as in GGA+U. Unfortunately, there are many equivalent ways to apply this correction (i.e. which choice of 'orbitals' to apply this to) yielding the same bandgap. Since defect levels are sensitive not to the bandgap itself, but to their position relative to the host band states, this ambiguity can result in many different predictions for defect ground states [20,21,32]. Furthermore, corrections for SIE are especially important for charged defect calculations as reducing interaction error tends to increase the ionicity of the crystal [33] resulting in more ionic relaxation as a defect state is populated and hence greater energy benefit for a charged defect.

Functionals which attempt to correct for SIE have recently emerged. One of the most robust and computationally tractable is the HSE functional [22]. The exchange-

correlation potential is divided into short and long-range components via a screening length parameter. For the long-range portion, things are unchanged from a typical GGA calculation. Short-range interactions have a portion of exact Hartree-Fock exchange mixed into the exchange potential, which partially corrects for SIE. This approach has been shown to greatly increase the accuracy of the bandgap as well as relative band positions for a wide variety of materials [34]. In this work, the amount of short-range HF-exchange mixing is set to 25% which is the amount suggested by the adiabatic connection theorem [35]. We use a screening length of .1 $^{-1}$ instead of the more typical .2 $^{-1}$ to account for the low dielectric strength, and hence poor screening of Zn_3P_2. This results in close agreement between our calculated bandgap and experiment (see Table 3.1).

3.3.2.1 Perturbation Extrapolation

However, the added memory requirements associated with the use of the HSE functional for a 2x2x2 supercell of Zn_3P_2 (320 atoms) make it computationally intractable with our existing resources (even smaller supercells represent a challenge). Here we have decided to use the 'perturbation extrapolation' method proposed by Lany, et al. [23] This is based on the idea of expressing the defect-influenced states in the basis of the states of the perfect bulk (assuming these form a complete basis):

$$\Psi_D(r) = \sum_{n,k} A_{n,k}\Psi_{n,k}(r) \tag{3.11}$$

If we model the bandgap correction of the HSE functional as a perturbation (H_p) of the perfect bulk system Hamiltonian (H_{bulk}) via a multiplier λ, the band energies shift:

$$e_{n,k}(\lambda) = \langle\Psi_{n,k}|H_{bulk}|\Psi_{n,k}\rangle + \lambda\langle\Psi_{n,k}|H_p|\Psi_{n,k}\rangle \tag{3.12}$$

$$e_{n,k}(\lambda) = e_{n,k}(0) + \lambda\frac{\partial e_{n,k}(\lambda)}{\partial\lambda} \tag{3.13}$$

If we apply this same perturbation to the defect system (with Hamiltonian $H_{bulk}+H_D$) then we have:

$$e_D(\lambda) = \langle \Psi_D | H_{bulk} + H_D | \Psi_D \rangle + \lambda \langle \Psi_D | H_p | \Psi_D \rangle \tag{3.14}$$

$$e_D(\lambda) = e_D(0) + \lambda \frac{\partial e_D(\lambda)}{\partial \lambda} \tag{3.15}$$

Within the framework of first-order perturbation theory and via Eq. (3.11) we have the final result:

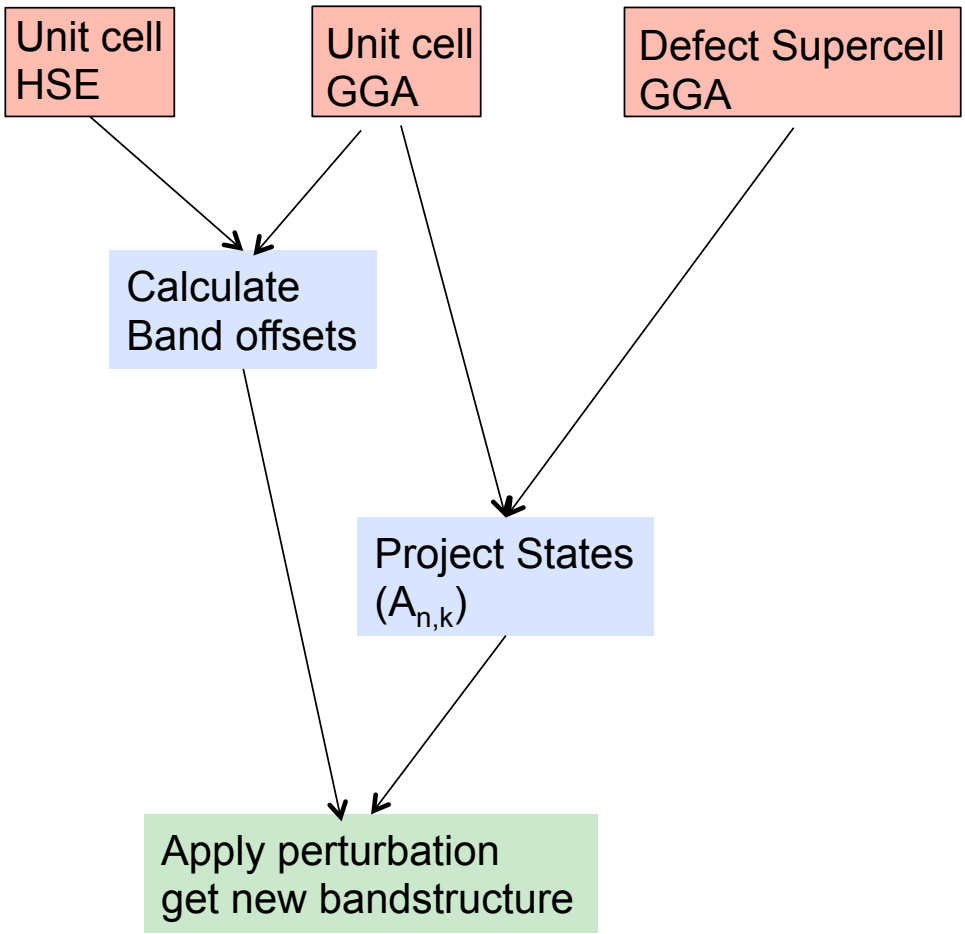
$$e_D(\lambda) = e_D(0) + \lambda \sum_{n,k} A_{n,k}^2 \frac{\partial e_{n,k}(\lambda)}{\partial \lambda} \tag{3.16}$$

As long as the assumptions of first-order perturbation theory are justified (i.e. unchanged wavefunctions upon application of the band-correcting perturbation H_p) we would predict the defect states to track the corrections to the perfect bulk states in proportion to the square of the coefficients in the expansion ($A_{n,k}^2$) of Eq. (3.11). As HSE doesn't significantly affect the dispersion of bands so much as their relative positions to each other, we expect that this approximation should be justified.

There is still a question as to whether the perfect bulk states form a reasonably complete basis to describe the defects. We are helped here by the fact that the static dielectric strength of Zn_3P_2 is low (~ 3) which leads to the tendency of the defects to not be very localized. The number of bands we need to calculate in the expansion of Eq. (3.11) is thus fairly limited.

The general workflow is shown in Figure 3.4. Perfect bulk unit cell wavefunctions are determined with GGA and HSE functionals. Then a 2x2x2 defect supercell is calculated only with the GGA functional. The defect supercell GGA wavefunctions are then projected onto the unit cell GGA wavefunctions obtaining $A_{n,k}$. Each band is then offset by the amounts prescribed by comparing the GGA and HSE unit cells. We can then make a prediction for what the supercell defect bandstructure would be

Figure 3.4: Perturbation Extrapolation Workflow - assuming that the perfect unit cell GGA forms a complete basis for the supercell defect wavefunctions, we project the GGA defect states onto the GGA perfect states and then use the offsets between the GGA and HSE unit cells to extrapolate HSE supercell behavior.



if we had been able to utilize the HSE functional.

3.3.2.2 Band offset determination

Determining the band offsets require that we determine the effect of the HSE functional on the bandstructure relative to GGA. Calculating the band shifts from GGA to HSE requires more than simple bulk calculations, as the band energies given by DFT are only referenced to the average electrostatic potential of the simulation cell. For a periodically repeated solid this is an ill-defined quantity which makes it impossible to directly compare band energies between cells with different constituents or for calculations performed with different functionals. However, if we have a region of equivalent potential (i.e. vacuum) in both GGA and HSE systems we can compare the bulk potentials in each system to this region, and subsequently compute the HSE and GGA band positions relative to each other. Care must be taken that the solid and vacuum regions are large enough that the electron wavefunctions become negligible in the vacuum and the effects of the surface states are not felt in the bulk. In this work, we use a 4 unit-cell slab along the non-polar [100] direction and an equal amount of vacuum (resulting in more than 35 of vacuum) - see Figure 3.5. Non-polar, non-reconstructed surfaces were used to avoid creating surface dipoles with the accompanying undesirable step in potential across the solid-vacuum interface.

3.4 Results

Zn_3P_2 exists in two phases, tetragonal and cubic. At room temperature and atmospheric pressure Zn_3P_2 forms a tetragonal phase with a 40-atom unit cell possessing symmetry P42/nmc. It is derived from the cubic fluorite structure with zinc at the center of a distorted phosphorus tetrahedra and the phosphorus surrounded by eight zinc sites lying roughly at the corners of a cube, only six of which are occupied [36] - see Figure 3.6.

Figure 3.5: Slab Geometry for band alignment determination. Red regions show areas where the potential was integrated in the bulk and vacuum regions.

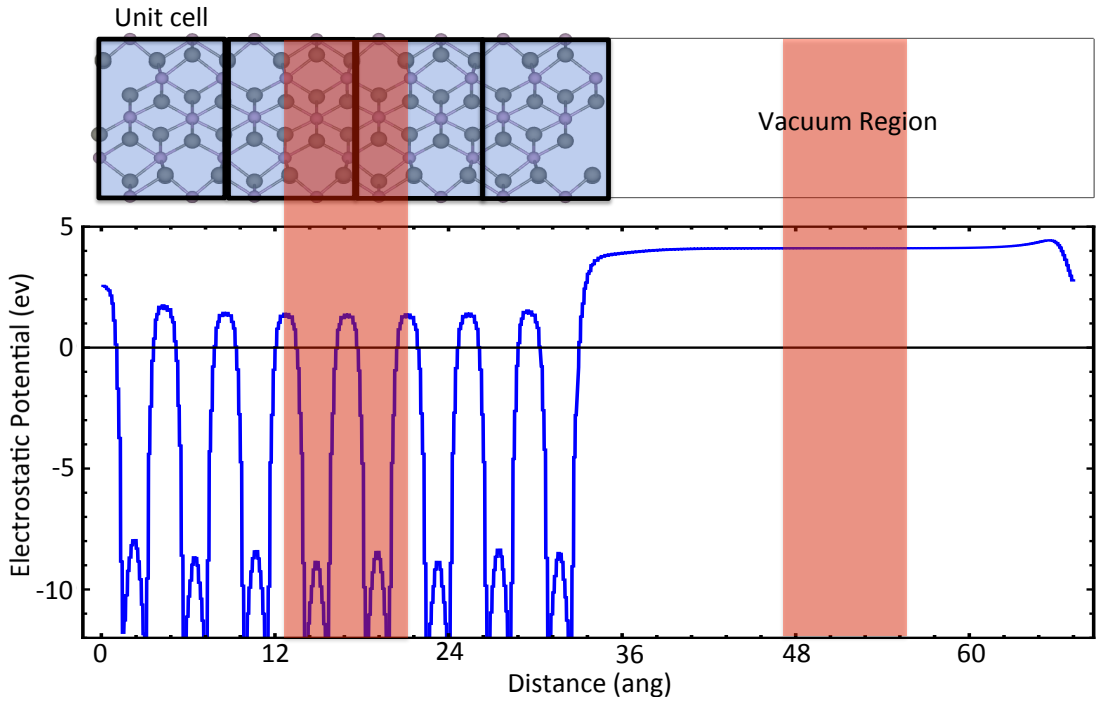
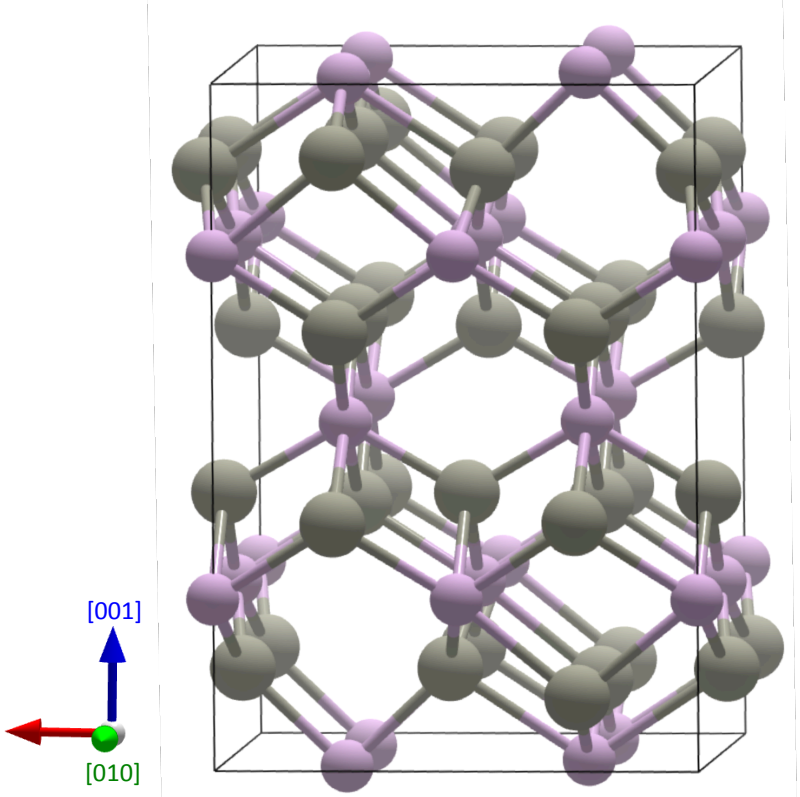


Figure 3.6: Tetragonal Zn_3P_2 has a 40-atom unit cell with P42/nmc symmetry.



We have investigated the intrinsic point defects of the room-temperature tetragonal phase. These defects include zinc vacancy (V_{Zn}), phosphorus vacancy (V_P), zinc interstitial (Zn_i), and phosporus interstitial (P_i). The antisite defects were also studied, but their formation energies are such that they would be unstable to dissociating into a vacancy-interstitial pair, thus they are excluded from the following discussion.

3.4.1 Computational details

All these calculations are performed within the Kohn-Sham framework (DFT) and utilize the projector augmented pseudopotentials as implemented in VASP [37–39]. We use valence configurations of 3d4s and 3s3p for zinc and phosphorus respectively with the Perdew, Burke, and Ernzerhof (PBE) potential [40]. GGA as well as hybrid functionals (HSE) for exchange-correlation were used, the specifics of which will be discussed later. For supercells of 2x2x2 unit cells (320 atoms) energies were calculated with two k-points (0,0,0) and (1/2,1/2,1/2) and a plane-wave basis with a cutoff of 300eV. These settings result in a numerical error on the order of 0.1 eV in the defect formation energy which is on par with the other sources of error (e.g. uncertainty in the bandgap and finite-size corrections). All calculations were performed considering spin polarization.

3.4.2 Structural Properties

The calculated structural and basic thermodynamic parameters are summarized in Table 3.1. Results for the lattice constant and formation enthalpy are in good agreement with experiment with both GGA and HSE functionals. However, the bandgap is severely underestimated with GGA, which at 0.32eV is only about 1/5 of the experimental result. Using the HSE functional, we find the bandgap to within 5% of experiment, highlighting the value of using HSE for this system.

Table 3.1: Calculated lattice constants, bandgap, and heat of formation of tetragonal Zn_3P_2 using both GGA-PBE and HSE functionals (using .1 $^{-1}$ screening length and 25% HF-exchange mixing).

functional	a (angs)	c/a	E_g (eV)	ΔH_f(eV)
GGA-PBE	8.108	1.408	.32	-1.79
HSE	8.160	1.390	1.42	-1.32
Experiment [16, 41]	8.097	1.286	1.49	-1.53

In order to accurately describe the defect levels we need to determine the effect of the HSE functional on the bandstructure (see Figure 3.7) relative to GGA. The alignment of GGA and HSE bandstructures is discussed in detail in Sec. 4 of the Appendix and is shown in Figure 3.8. The lowering of the VBM with HSE is due to the reduced self-interaction of the phosphorus p-orbitals which primarily form the highest-energy valence bands. The upward shift of the CBM is due to the reduced hybridization of the Zn-s and P-s orbitals that make up the lower part of the conduction band with the valence band states. With this alignment calculated we can compute the GGA-to-HSE band offsets needed for the 'perturbation extrapolation' method.

3.4.3 Intrinsic defect formation energetics

We are now in the position to apply all the supercell corrections listed in Section 2.3 above. We turn our attention to the predicted stability and doping effects of the various intrinsic defects. Here it is useful to make a few notes about the following discussion; first, we may find that a defect level lies outside of the bandgap region predicting that this charged state is unstable as a defect localized state. However, it may still be possible to bind electrons to the defect through electrostatic interation and form hydrogenic effective-mass states just within the band edge. A discussion of these states is not in the scope of this work. Furthermore, we ignore the effects of both formation volume and formation entropy in computing the defect formation energies. Formation volume is directly related to the change in volume when a defect

Figure 3.7: DFT Bandstructure - calculated with HSE functional. Partial density of states shows makeup of VBM mainly of phosphorus p-character, while the CBM is mixed p and s-character from phosphorus and zinc states.

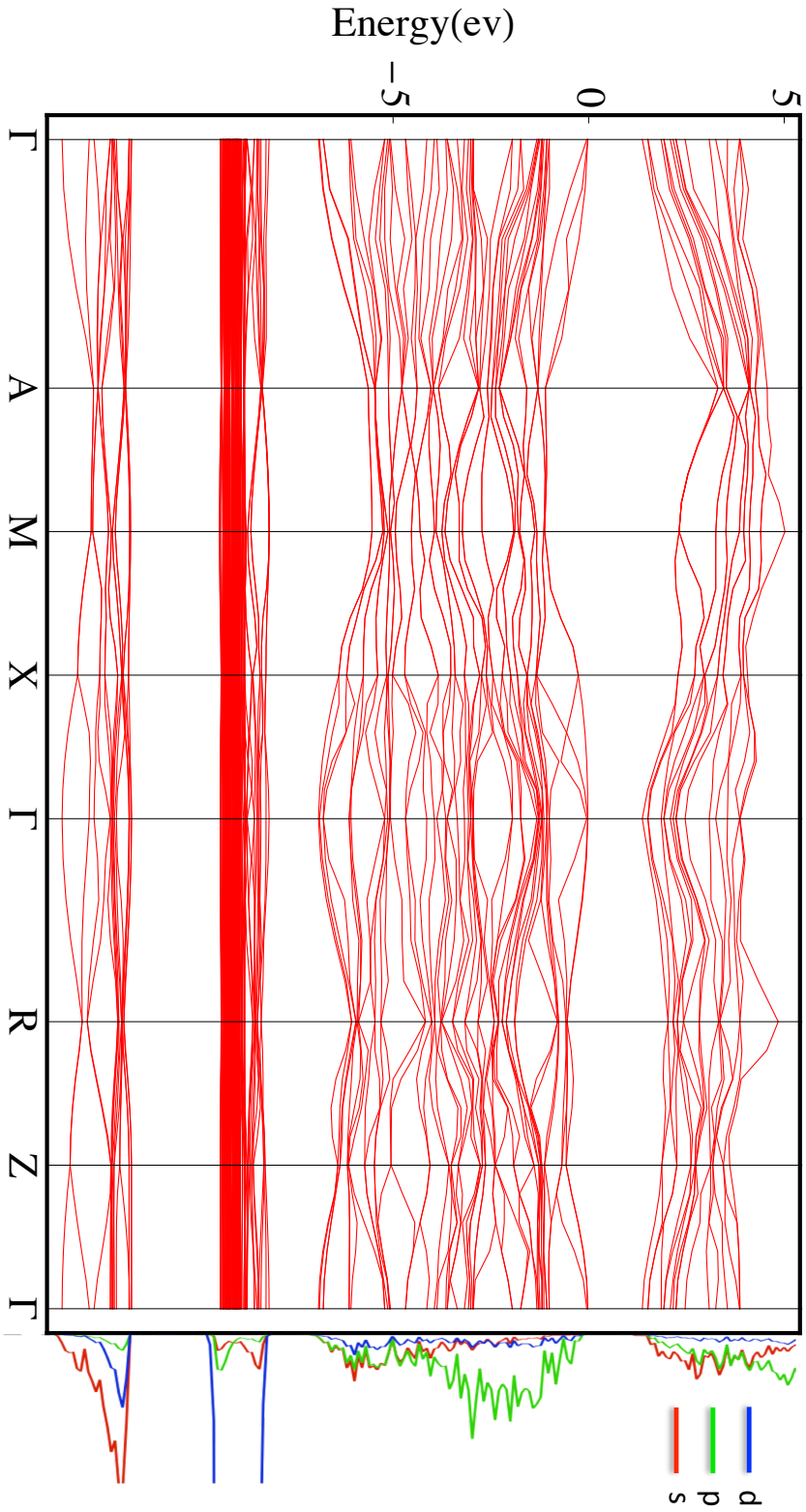
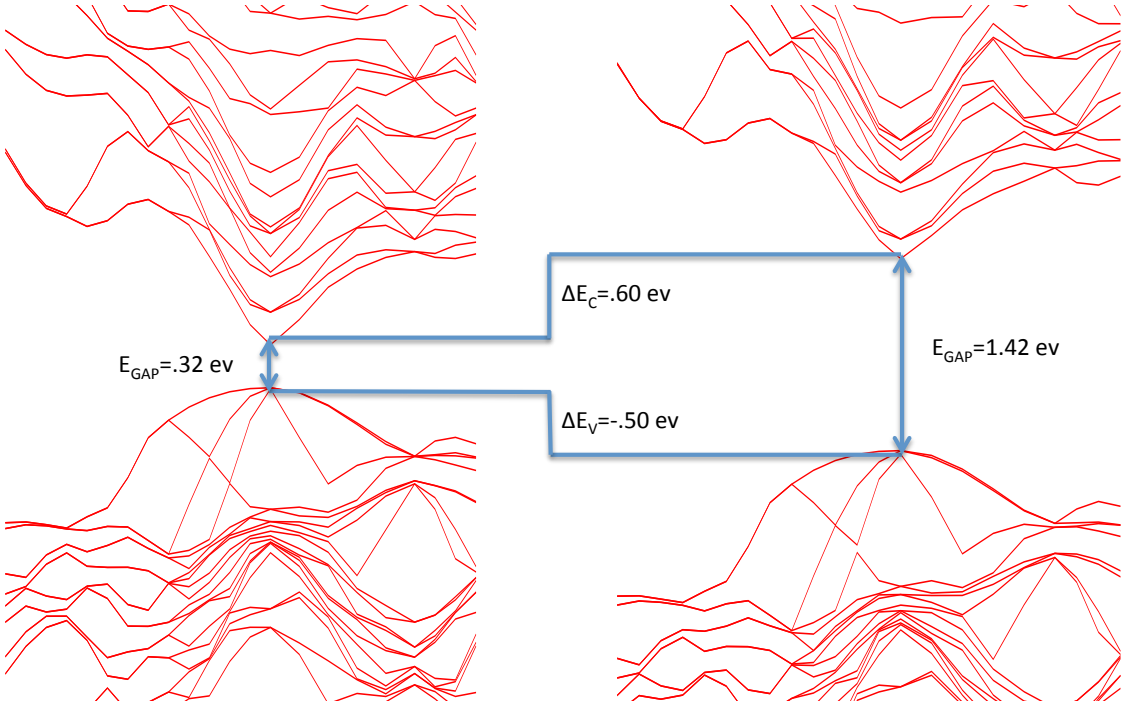


Figure 3.8: GGA to HSE Band Alignment. Bandgaps for GGA (left) and HSE (right) are labeled at the gamma point. Calculated offsets for the VBM and CBM are shown.



is created, but is typically only important for degenerate defect regime or for very high pressures. Formation entropy for point defects is typically on the order of a few k_B so are only important for very high temperatures - additionally, the entropy of point defects in the same system tend to be similar and so largely cancel when we compare the likelihood of one defect over another.

3.4.3.1 Zinc Vacancy

Zinc is surrounded by four nearly equidistant, tetragonally coordinated phosphorus atoms. Removing a zinc atom from the lattice leaves four dangling bonds from the neighboring phosphorus atoms with mainly p-character - see Figure 3.9. These bonds are occupied with only six electrons, forming three low energy and one empty higher-energy defect state. This empty bonding state can capture electrons and form an acceptor defect with a -1 or -2 charge. As the defect states have a strong valence band character (e.g. in terms of symmetry), we would expect shallow defect behavior.

It is instructive to study the structural relaxations around the V_{Zn} site as this is intimately related to the occupancy and energetics of the defect state. Removing a zinc atom to form a neutral defect causes the four neighbouring phosphorus atoms to relax away from the defect site as they seek to maximize their bond overlap with the remaining zinc atoms in the lattice. However, this movement also shifts the dangling bonds in the vacancy region higher in energy, fighting this relaxation. The phosphorus atom closest to the defect site (the P-atom to the 'north' of the vacancy site in Figure 3.9) has more of its density in the vacancy region and subsequently relaxes the least. As the defect becomes occupied (as in the -1 or -2 charged states) the electrons are chiefly populating around the phosphorus ions which cause a further relaxation away from the defect due to increased coulombic repulsion. As the relaxation is relatively slight in all defect charge states, we would expect that the defect itself would be delocalized. This is born out as the charge density is poorly screened and well dispersed throughout the lattice.

Getting the correct energetics and occupancy of the defect states is primarily important for their effect on the formation energy of the defects, which is our central goal. Looking at the formation energy plots in Figure 3.10, zinc vacancies exist as charged defects for all but very p-type regimes (fermi level close to VBM). The small energy difference between the different charged states for a fermi level at the VBM can be expected from the small difference in lattice relaxations associated with the charged states of V_{Zn}. The formation energies in the n-type regime are low enough in the Zn-deficient growth environment for this to be an important defect.

Figure 3.9: V_{Zn} Electronic Structure - Partial Density of States of the neutral defect for the four nearest neighboring phosphorus atoms. Defect states highlighted in yellow, show mainly p-character. Bond lengths are given relative to distance to defect site in perfect bulk cell.

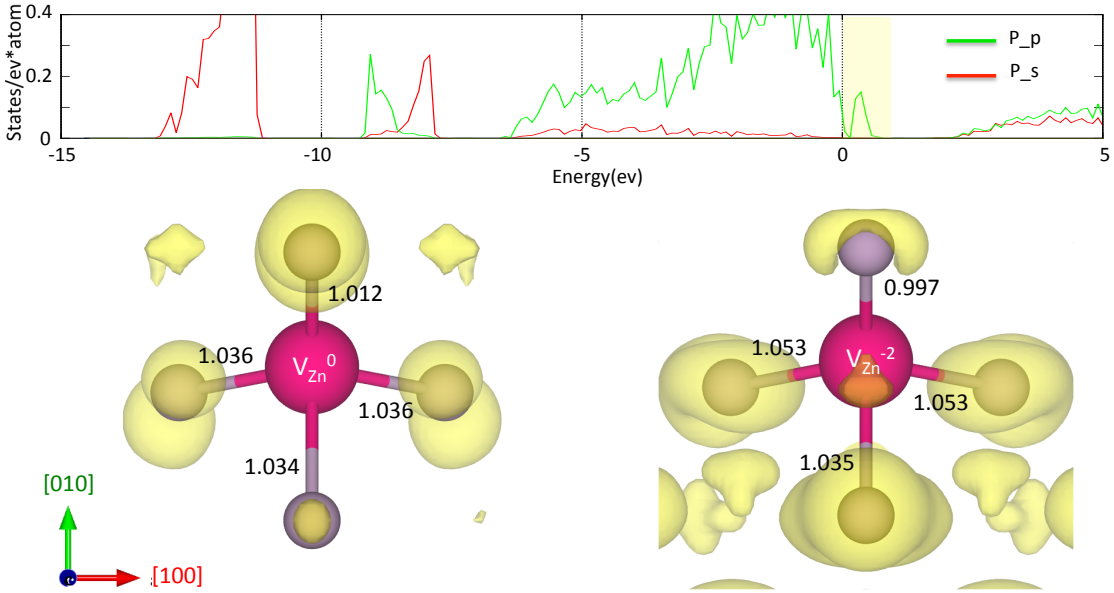
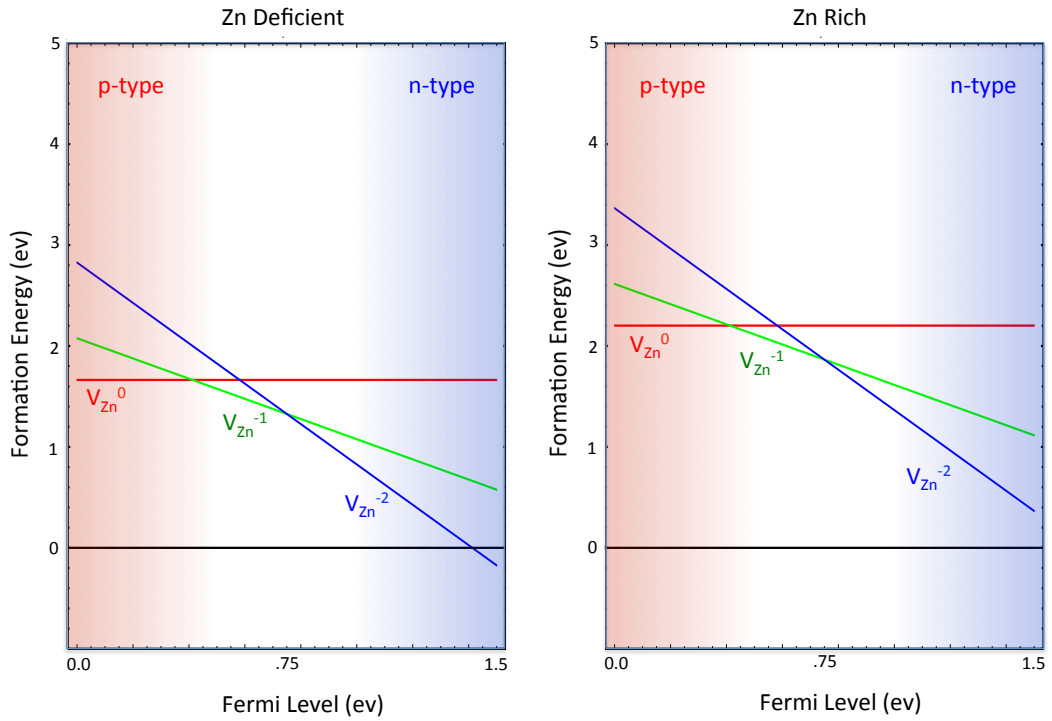


Figure 3.10: V_{Zn} Formation Energy. Acceptor-type defect, plotted for the 0,-1,-2 charged states as the fermi level varies from the top of the VBM (p-type) to the bottom of the CBM (n-type). Favorable growth condition on the left.



3.4.3.2 Phosphorus Vacancy

Phosphorus is surrounded by six nearest neighbor zinc sites, four being almost equidistant and two zinc atoms about 15% further away. Phosphorus normally exists in a -3 oxidation state and when we remove a phosphorus atom the electrons from the surrounding zinc atoms have only high energy bonding states to move into with mainly Zn-p and Zn-d character - see Figure 3.11. We would expect that these states would want to depopulate and form +1 and +2 charged defects as these electrons are no longer needed for bonding a phosphorus atom to the lattice. The symmetry of the defect states is not similar to the conduction band symmetry so we would expect deeper defect behavior than for V_{Zn}.

Removing a phosphorus atom results in a defect state that is effectively screened by the neighboring zinc atoms and highly localized between the nearest zinc sites. The high degree of localization results in a large lattice relaxation, especially for the two zinc atoms closest to the defect site. Since the majority of the electron density is concentrated in the defect region, the four closest zinc atoms relax closer to each other in order to lower the energy of the occupied defect localized states. The two furthest zinc atoms relax away from the defect region due to increased coulombic repulsion from the similarly charged zinc atoms closer to the defect. As the defect becomes depopulated there is a slight outward relaxation since the benefit in lowering the energy of the defect states is reduced and the zinc atoms want to increase their bonding with the rest of the lattice. The high degree of localization would also suggest deep defect behavior.

Looking at the formation energies in Figure 3.12, we see deep defect behavior where the defect only becomes charged for fermi levels in the neutral to p-type regime. Even for favorable growth conditions these defects are too high in energy to play a significant role in the dopability of zinc phosphide.

Figure 3.11: V_P Electronic Structure - Partial Density of States of the +2 charged defect for the nearest neighboring zinc atoms. Defect states highlighted in yellow, show mainly Zn-p and Zn-d character. Bond lengths are given relative to distance to defect site in perfect bulk cell.

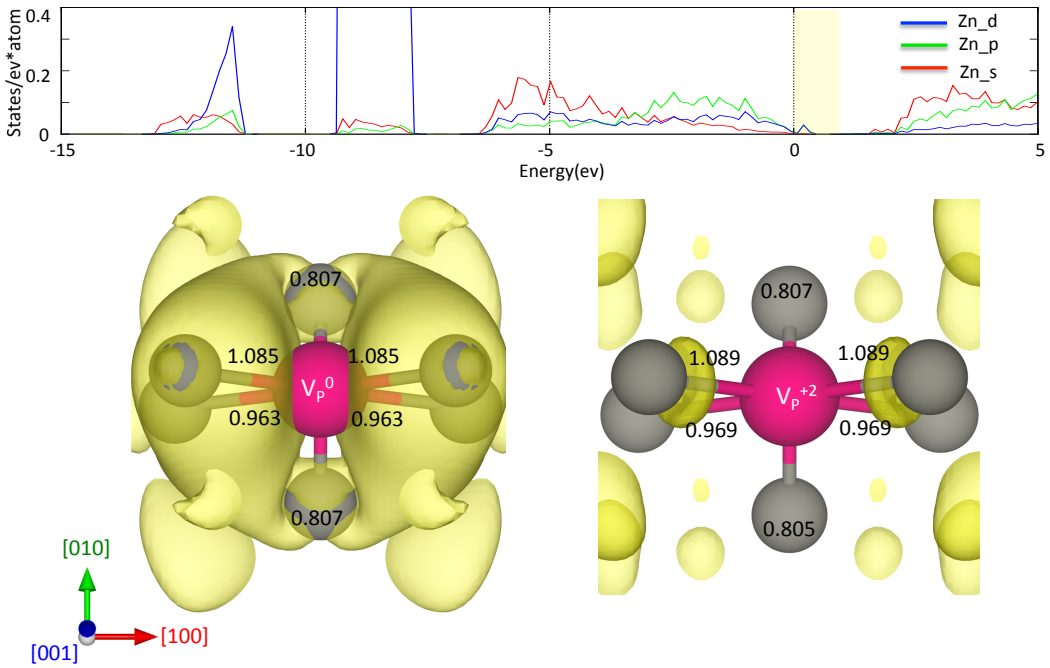
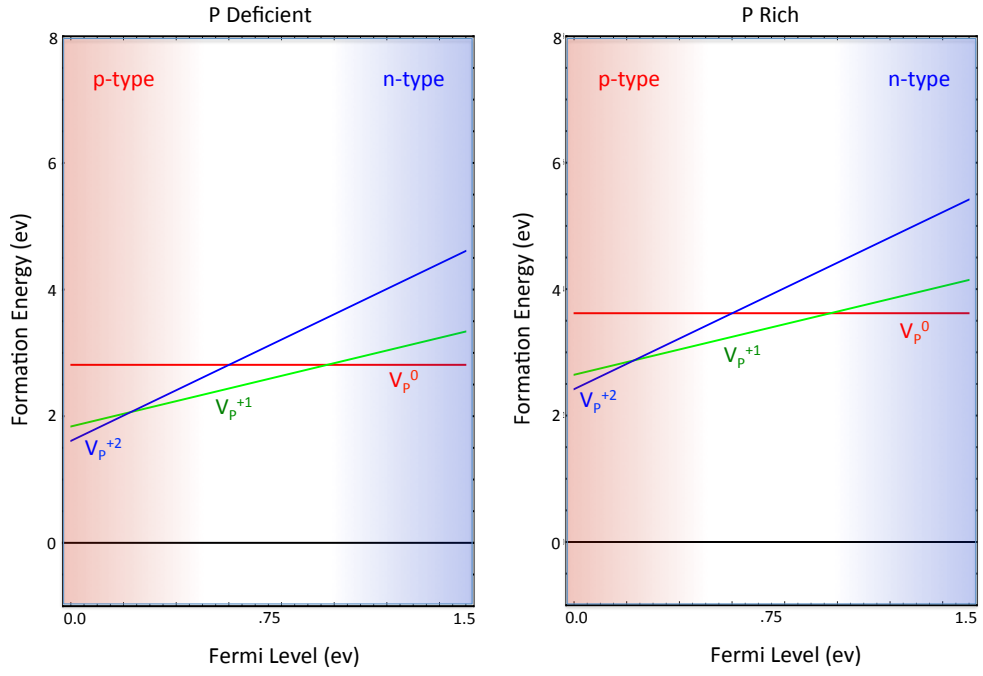


Figure 3.12: V_P Formation Energy. Donor-type defect, plotted for the 0,+1,+2 charged states as the fermi level varies from the top of the VBM (p-type) to the bottom of the CBM (n-type). Favorable growth condition on the left.



3.4.3.3 Zinc Interstitial

For both zinc and phosphorus interstitials we voxelized the Zn_3P_2 unit cell and tested the sites in order of furthest distance from neighboring atoms. Zinc interstitials are predicted to form in the voids on the zinc plane of atoms tetrahedrally coordinated with four phosphorus atoms and in-line with zinc atoms in adjacent planes. Incorporating an extra zinc atom into the lattice forms a localized defect state of mainly ionic Zn-s character since the neighboring phosphorus atoms have closed-shell configurations. As for zinc atoms in the perfect bulk, the interstitial zinc has a tendency to depopulate the s-shell states. We would expect donor defect behavior with a $+1$ or $+2$ charged state and this is what our calculations show. The symmetry of the defect states are similar to the conduction band character, consequently we would expect shallow behavior.

The neighboring zinc atoms are affected the most by the incorporation of the Zn_i defect as they relax away from the similarly charged interstitial ion. The phosphorus atoms relax closer to the interstitial site in general as there has been a partial charge transfer from the Zn_i defect. There is only minor differences in relaxation between the various charged states and the defect is poorly screened with a delocalized wavefunction centered on the majority of the zinc atoms in the lattice. All of which would suggest shallow defect formation.

In Figure 3.14, the formation energy plots show shallow behavior where the defect becomes charged for all but fermi levels high in the n-type regime. As the fermi level drops to the VBM the formation energy of the defects becomes very small for the favorable growth conditions (Zn-rich regime). These defects should be important to consider.

Figure 3.13: Zn_i Electronic Structure - Partial Density of States of the -2 charged defect for the interstitial Zn site. Defect states highlighted in yellow, show mainly Zn-s character. Bond lengths are given relative to distance to defect site in perfect bulk cell.

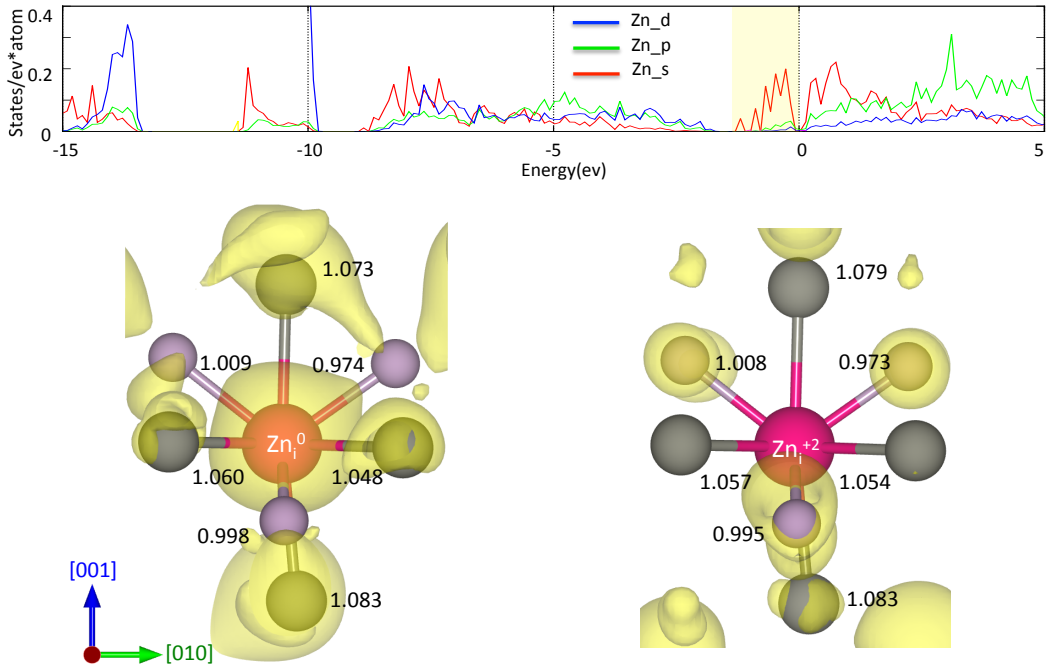
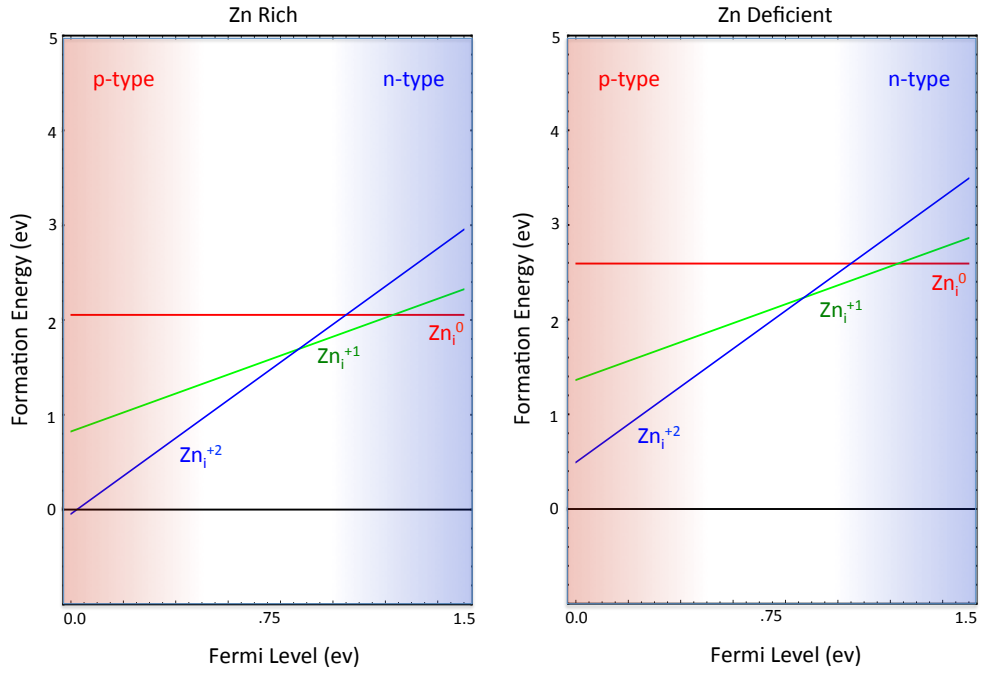


Figure 3.14: Zn_i Formation Energy. Donor-type defect, plotted for the 0,+1,+2 charged states as the fermi level varies from the top of the VBM (p-type) to the bottom of the CBM (n-type). Favorable growth condition on the left.



3.4.3.4 Phosphorus Interstitial

The most energetically favorable position for the phosphorus interstitials are in the voids in the zinc plane of atoms nearly equidistant from three zinc atoms and directly above a phosphorus atom displaced from the plane below. The additional electrons form states of mainly covalent p-character between the P_i site and neighboring zinc and phosphorus atoms - see Figure 3.15. Due to the natural oxidation state of phosphorus (-3) we expect that the defect should be an acceptor type, which is confirmed by our calculations. Since the symmetry of the defect states is similar to the VBM, we would predict shallow defect behavior.

The neutral defect state is highly localized, so P_i causes relatively large ionic relaxation away from the defect site, though for the two zinc atoms most involved with the covalent bonding this relaxation is reduced because this movement also increases the energy of the populated defect states. The charged state is poorly screened and very delocalized across the lattice. Consequently, there is very little relaxation as the defect captures electrons from the conduction band. Apart from energetics, the severe delocalization of the charged state and good symmetry match between the defect and valence states causes almost band-like behavior where we should readily capture electrons from the conduction band.

The neutral defect has relatively low formation energy for P-rich growth conditions. Due to the small differences in relaxation between the charged defect states, there is only a small difference in formation energy between them at the VBM. Thus, for fermi levels even modestly into the n-type regime the formation energy of the charged defects becomes very small making these critical defects to consider.

Figure 3.15: P_i Electronic Structure - Partial Density of States of the neutral defect for the interstitial phosphorus site. Defect states highlighted in yellow, show mainly P-p character. Bond lengths are given relative to distance to defect site in perfect bulk cell.

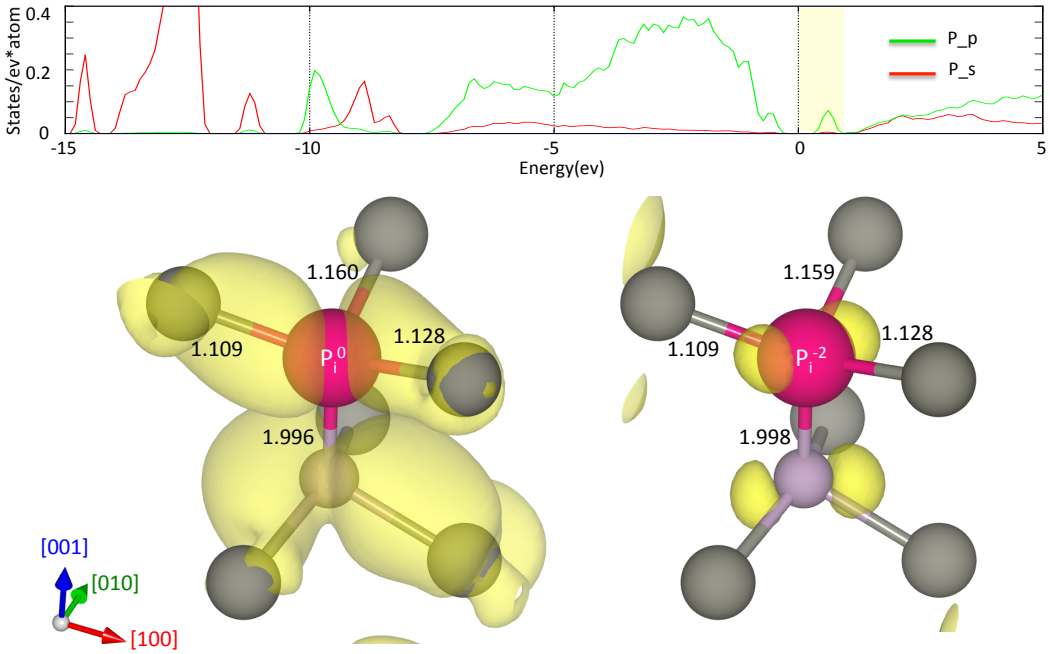
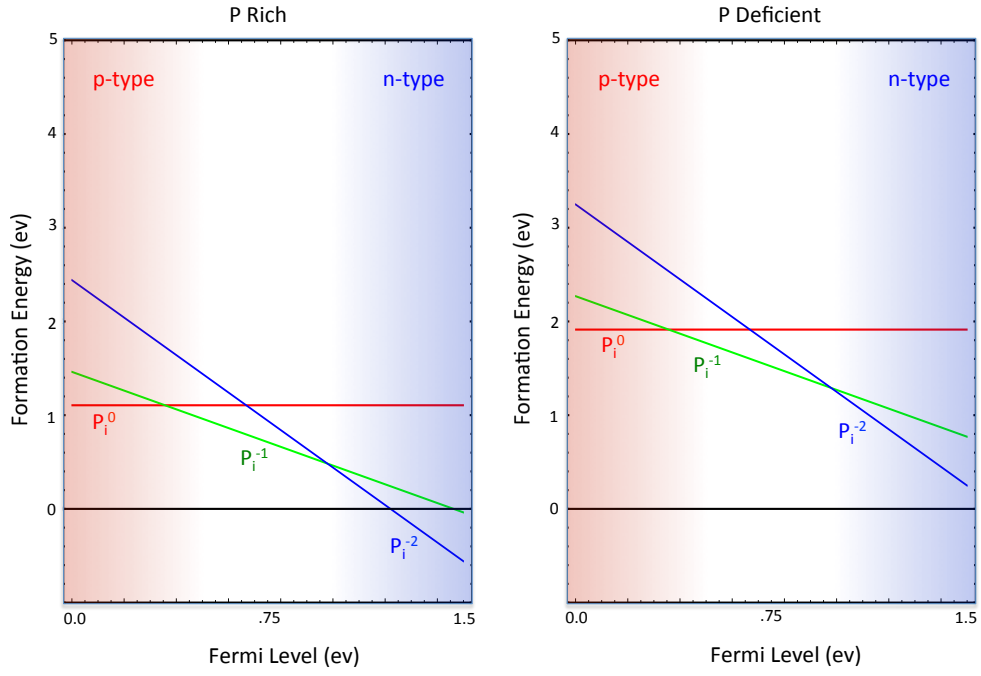


Figure 3.16: P_i Formation Energy. Acceptor-type defect, plotted for the 0,-1,-2 charged states as the fermi level varies from the top of the VBM (p-type) to the bottom of the CBM (n-type). Favorable growth condition on the left.



3.5 Conclusion

From our calculations, the likely candidate for the lack of n-type dopability is the P_i defects. Both of the acceptor defects (V_{Zn} and P_i) have low formation energies as we move the system into the n-type regime. While there are more suitable V_{Zn} sites, the P_i defect are significantly less costly and should be the vastly more prevalent defect. The Zn_i defects would tend to aid n-type doping as they are electron donors, though they have too high a formation energy for fermi levels even moderately n-typed to significantly compensate for the P_i defects.

Thus, as we dope our system with electrons we create a large amount of acceptor defects which act as 'electron-sinks' and capture mobile electrons from the conduction band and neutralize the doping. In fact, the P_i defect requires zero formation energy for fermi-levels midway towards the conduction band. This would 'pin' the fermi-level - that is, as we try to approach this level of doping a massive amount of P_i defects would form in the crystal. Since the creation of P_i defects fight n-type doping, reaching this level would not be expected. There is some hope in that we would anticipate the P_i defects to repel each other (especially the highly charged state), thus limiting their concentration, and some n-type doping would then survive. Indeed, initial results show a significant penalty in having two fully charged P_i defects within the same 2x2x2 supercell. However, this still corresponds to a large defect concentration and there is low likelihood that the desireable properties of Zn_3P_2 such as high minority carrier diffusion lengths would survive in the regime where P_i defects saturate.

Solving for the fermi level self-consistently using the above formation energy functions of the various defects, we predict an intrinsic fermi-level of 0.55 eV, which is mildly p-type. Pushing the fermi level beyond roughly 1.05 eV becomes impossible as the number of defects exceeds the numer of sites in the cell at this point.

Since the P_i defects are the problematic defects, any means to suppress them

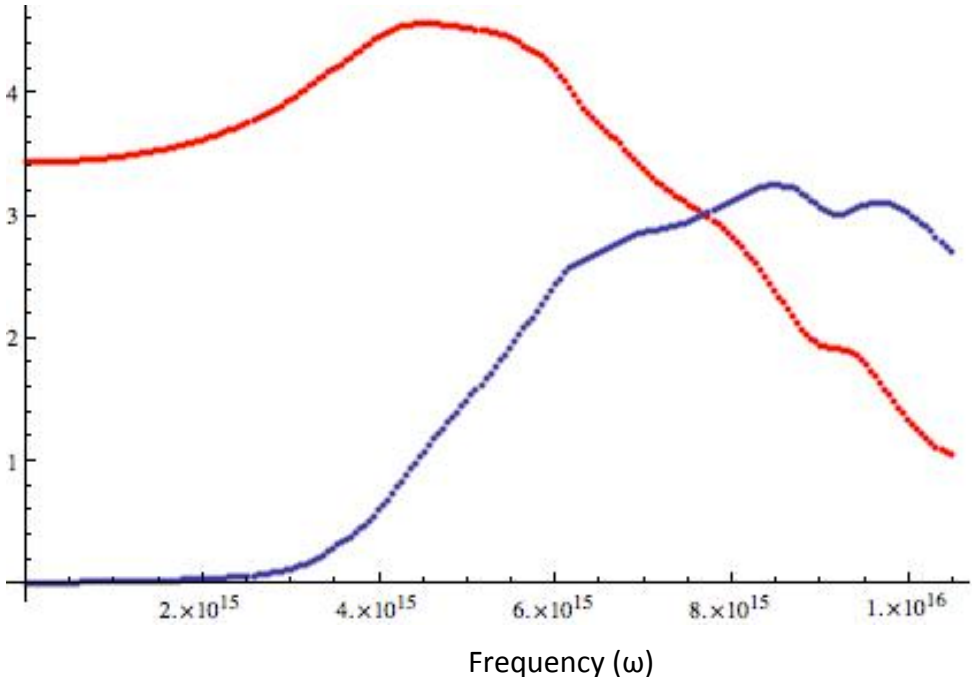
should help the n-type doping issue. For extreme Zn-rich growth conditions, where the acceptor P_i defects are suppressed and the donor Zn_i are enhanced there may be some hope of weakly n-type materials being formed. A suppression of interstitial defects in general, such as straining the crystal as it is grown may prove fruitful. Cluster doping [42] donor atoms with ones that suppress P_i formation would be another avenue to explore.

In closing, we have presented a state-of-the-art approach in performing defect calculations using DFT. Great care has to be taken in correcting for finite-size effects inherent in the supercell formalism, which is especially problematic for low dielectric and low symmetry systems like zinc phosphide. Correct formation energies also depend greatly on an accurate picture of the defect influence on the band structure. This involves a careful choice of the exchange-correlation potential to use, often restricted by the computational resources available. Here we have implemented, for the first time, a perturbative method to extend results of a highly accurate functional on a small unit cell to a much larger collection of atoms. The results are promising and the method deserves further investigation and refinement.

3.6 Optical absorption addendum

Though it is a small digression of the rest of the work in this chapter, because of its importance to future efforts, it is useful to quickly discuss the current progress in determining the optical properties of this system. In the 'big-picture', this zinc phosphide task is part of a broader effort to find low-cost materials to use as photovoltaics, not all of which need even have been created yet. As a practical matter it would be very helpful if we only perform the expensive defect/doping calculations for materials that have a favorable optical absorption profile for use in collecting solar radiation. Here we need a very accurate bandstructure, even for excited states. As mentioned in the previous chapter, DFT makes no guarantees about the energetics of any state aside from the highest occupied orbital (Sec. 2.4.2.2). While the 'guesses' it makes

Figure 3.17: Zinc phosphide optical response generated from G_0W_0 method. Red curve is refractive index (n) and blue curve is extinction coefficient (k). Both compare favorably to recent experimental measurement [16].



about the other ground states and the excited states are often reasonable, for optical calculations it makes sense to use a theory like GW (Sec. 2.5) which should at least converge to the physically correct results. We have worked on applying this method to this system (and others) and have gotten some encouraging results - see Fig(3.17). From even the lowest hanging fruit in the theory (G_0W_0) which was within about an order of magnitude as fast to run as DFT, we have achieved remarkable agreement with experimental measurements for the frequency-dependent index of refraction and extinction coefficent. It is clear that, especially as computers get more capable, that the GW method has a place in screening candidate photovoltaic materials. With more research, this method may even have uses in the more mundane world of defect energetics as part of the perturbation extrapolation procedure.

Chapter 4

Multiphase equation of state of tantalum from first-principles

In this chapter we are concerned with constructing an equation of state (EOS) over a wide range of temperatures and pressures. The typical means of investigating thermodynamic properties is through 'classical' molecular dynamics (MD) where the atomic motion is calculated from Newtonian mechanics with the electronic effects abstracted away into an interatomic potential function. An *ab initio* approach is useful here as a classical potential is typically valid for only portions of the phase diagram [43, 44] (i.e. whichever part it has been fit to). Furthermore, for extremes of temperature and pressure quantum effects become critical to accurately calculate the EOS and are very hard to capture in even complex model potentials. As in the previous chapter, this necessitates extending DFT into uncommonly tread territory. Here the challenge is in predicting the finite temperature response of our system with a theory that is inherently zero K. Additionally, we must cover a large area of phase space efficiently and with enough accuracy to faithfully capture phase coexistence regions which occur at the limits of typical DFT validity [45]. Our end result is an EOS for Ta which covers an unprecedented range of pressures and temperatures with a high degree of accuracy.

4.1 Introduction

The accurate determination of high pressure and temperature equations of state is fundamental in a variety of fields. For instance, geophysics and planetary sciences typically require knowledge of material responses in extreme conditions [46]. There is also an ongoing and increasing interest in modelling high-velocity impacts via multiscale methods [47, 48] which also demand equations of state over a wide range of conditions. In addition, many of these applications require quantities that are dependent on high-order derivatives of the free energy (such as the speed of sound) demanding that the EOS is very smooth as well as wide-ranging.

Here we present a comprehensive calculation of the equation of state for tantalum. Our approach builds upon the previous investigations of Ta [46,49,50] in a few major ways:

1. It is based on direct *ab initio* calculations, without relying on force-field fitting, thus enabling us to assess the quality of these force fields.
2. It covers all the phases - solid, liquid, gas and all associated two-phase equilibria.
3. It includes the effects of electronic excitations.
4. We deliver not only the phase-boundaries, but also the free energies of all the phases which is a requirement for multi-scale modelling applications.

We deliver the EOS in the form of the Helmholtz free energy as a 2D function of volume and temperature (V,T) phase space. Densities span from high pressure solids to well beyond the critical point with a pressure range of 0-7 Mbars and temperatures from 0-20,000K.

In order to investigate material behaviour across all the various phases we resort to the unifying approach of thermodynamic integration (TI). We utilize *ab initio* molecular dynamics (AIMD) to determine properties at a fixed density and temperature over a (V,T) grid covering a large amount of phase space. Using TI we can

then employ analytical statistical mechanics models both as a reference point to the integration as well as a way to extrapolate our results beyond the directly calculated region. Taking advantage of these models also helps reduce the required density of expensive *ab initio* generated points. These also have the nice property that local deviations from models are easily parameterized by simple, physics neutral, functional forms (ie. splines).

As a final requirement for this project, we would like to be able to calculate various transport properties, such as viscosity, from the same *ab initio* data as is used for the EOS. A key insight here is provided by the fluctuation-dissipation theorem which asserts that random forces and frictional forces result from the same microscopic mechanism. This results in the Green-Kubo relations discussed below.

4.2 Ab-initio molecular dynamics

Of primary concern with any molecular dynamics methods is how to describe (i.e. approximate) the interatomic interactions. The level of approximation used in classical MD is to abstract away the electronic degrees of freedom into a predetermined potential function. This potential is generally fit to either empirical data or electronic structure calculations. Typically, the complex interactions of the physical system are segmented into two-body, three-body, and many-body contributions as well as short and long-range terms. Though the classical approach has had many successes [51–53], the requirement of using a predefined potential comes with significant drawbacks. Some of the most serious occur in systems where either many different atom or molecule types lead to a runaway cascade of interactions that need to be parametrized as well as systems where the electronic structure changes qualitatively over the course of the simulation, perhaps invalidating the model potential's assumptions. [54]

With the advent of *ab intio* MD [55,56] we can greatly extend the reach of classical MD. This approach unifies Newton's and Schrödinger's equations and allows for complex systems to be simulated without the need for adjustable parameters. Forces acting on the nuclei are computed on-the-fly from electronic structure calculations that are performed dynamically with ionic motion. In this way, electronic degrees of freedom are preserved and not simply integrated out beforehand. While the added computational requirements mean that the relaxation times and correlation lengths that we can probe with *ab inito* MD are much smaller than with standard MD, unanticipated atomic interactions can now occur giving this approach the ability to anticipate new phenomena. As compared with classical MD, the level of approximation has been shifted from the choice of model interatomic potential to the particular approximation used to solve the Schrödinger equation.

In AIMD the interatomic forces are determined from the potential energy surface as:

$$F_I = -\nabla_R \Phi(R) \tag{4.1}$$

where R is the nuclear coordinates and $\Phi(R)$ is the potential energy function. Here we have already implicitly assumed the Born-Oppenheimer approximation where the electrons are in an instantaneous equilibrium with the much heavier nuclei so that the electronic subsystem can be treated independently with the nuclear positions frozen at a constant R. The potential function $\Phi(R)$ can be defined as:

$$\Phi(R) = \langle \psi_0 | \hat{H}_e(r_i; R) | \psi_0 \rangle + E_{II}(R) \tag{4.2}$$

where $E_{II}(R)$ is the Coulomb energy of the ions and $\hat{H}_e(r_i; R)$ is the electronic many-body hamiltonian which depends on electronic coordinates r_i and parametrically on R (as in Eq. 2.6). The forces do not depend on any adjustable parameter, but only on R which gives AIMD its predictive power.

However, finding the ground state ψ_0 of the electronic hamiltonian involves solving the many-body Schrödinger equation which is a high-dimensional eigenvalue problem with a significant associated cost. If we consider a real-space grid, where each coordinate is coarsely divided into as few as 100 points, for N electrons, a minimum of 10^{6N} grid points are required. The solution of a single Ar atom would require more grid points than the number of electrons in the universe. Fortunately, this 'dimensionality bottleneck' can be largely solved by considering the electrons via a mean-field approach as in DFT (Sec. 2.4) which works with electron density $\rho(r)$ (3 dimensional) rather than a 3N-dimensional wavefunction as its primary input. The potential energy is then calculated as:

$$\Phi(R) = E_{KS}[\{\psi_i[\rho(r)]\}; R] + E_{II}(R) = E[\{\psi_i\}; R] \tag{4.3}$$

where E_{KS} is the Kohn-Sham energy (see Eq. 2.26) and ψ_i are the single-particle orbitals inherent to the Kohn-Sham formalism.

There are two general approaches to AIMD, Born-Oppenheimer and Car-Parrinello, both of which are briefly described below.

4.2.1 Born-Oppenheimer MD

In Born-Oppenheimer MD (BOMD) [56, 57] the electronic structure is fully relaxed to its ground state at every MD time step. This has the benefit of being able to solve for the electrons in a time-independent fashion (i.e. the location of the nuclei are considered to be fixed). Once the electronic ground state is found we can generate the forces on the nuclei (via Eq. 4.1) and propagate them according to Newtonian mechanics. Considering that DFT is being utilized to solve for the electronic structure, we develop the equations of motion (EOM) for a one-particle effective hamiltonian where the potential energy is minimized with respect to $\{\psi_i\}$ subject to the constraint that the single-particle orbitals are orthonormal $\langle\psi_i(r)|\psi_j(r)\rangle = \delta_{ij}$; leading to the

following Lagrangian:

$$\mathcal{L}_{BO} = \frac{1}{2}\sum_{I=1}^{N} M_I \dot{R}_I^2 - min_{\{\psi_i\}} E[\{\psi_i\}; R] + \sum_{i,j} \Lambda_{ij}(\langle\psi_i|\psi_j\rangle - \delta_{ij}) \tag{4.4}$$

where Λ_{ij} are the respective Lagrangian multipliers. Through the corresponding Euler-Lagrange equations:

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{R}_I} = \frac{\partial \mathcal{L}}{\partial R_I} \tag{4.5}$$

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \langle\dot{\Psi}_I|} = \frac{\partial \mathcal{L}}{\partial \langle\Psi_I|} \tag{4.6}$$

we can obtain the equations of motion (EOM):

$$M_I \ddot{R}_I = -\frac{\partial E}{\partial R_I} + \sum_{i,j} \Lambda_{ij} \frac{\partial}{\partial R_I}\langle\psi_i|\psi_j\rangle - 2\sum_i \frac{\partial\langle\psi_i|}{\partial R_I}\left[\frac{\delta E}{\delta\langle\psi_i|} - \sum_j \Lambda_{ij}|\psi_j\rangle\right] \tag{4.7}$$

$$0 \lesssim -\hat{H}_e\langle\psi_i| + \sum_j \Lambda_{ij}|\psi_j\rangle \tag{4.8}$$

where the first term on the right of Eq.(4.7) is the Hellmann-Feynman force (see 'Phonon Calculation' below). The second term is the Pulay force due to the orthonormality constraint and vanishes only if the basis functions used to represent ψ_i do not depend on R (e.g. for a plane-wave basis this will vanish). The final term is a consequence of the implicit dependence on atomic positions through the basis expansion coefficient regardless of the basis set used. This term vanishes as long as ψ_i is an eigenfunction of the hamiltonian. In practice, this is only achieved at exact self-consistency which can never be assumed, so this is never zero (and also leads to the inequality of Eq. 4.8). As a result, forces are linearly dependant on the error in charge density, which is why it is much harder to calculate accurate forces than energies (if electron density is known to linear order then the energy is known to cubic order via the '2n+1' theorem).

Since there is no electronic dynamics in this approach (we are always at the electronic ground state at every moment) the integration time-step is solely determined by the time scale of the nuclear motion. This is the great advantage of this technique as this is orders of magnitude larger than for the electronic motion in the system. However, the main drawback of this approach is the large expense in finding the electronic ground state at every time step.

4.2.2 Car-Parrinello MD

The central idea of Car-Parrinello MD (CPMD) [58] lies in exploiting the time-scale separation between the fast electrons and far slower nuclear motion (as in BOMD), as well as taking advantage of the smooth evolution of the electronic system. Here we decompose the quantum mechanical system into two classical systems with separate energy scales (i.e. *both* electrons and nuclei evolve dynamically). With CPMD the electronic degrees of freedom are incorporated into the Lagrangian in a classical manner:

$$\mathcal{L}_{CP} = \frac{1}{2}\mu \sum_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle + \frac{1}{2}\sum_{I=1}^{N} M_I \dot{R}_I^2 - E[\{\psi_i\}; R] + \sum_{i,j} \Lambda_{ij}(\langle \psi_i | \psi_j \rangle - \delta_{ij}) \qquad (4.9)$$

where we have introduced a fictitious inertia to the electronic degrees of freedom via a mass parameter μ. It should also be highlighted that we are only concerned with the instantaneous potential energy of the system - this should be compared to Eq.(4.4). Again, applying the Euler-Lagrange equations of Eqs.(4.5, 4.6) we obtain the CP equations of motion:

$$M_I \ddot{R}_I = -\frac{\partial E}{\partial R_I} + \sum_{i,j} \Lambda_{ij} \frac{\partial}{\partial R_I} \langle \psi_i | \psi_j \rangle \qquad (4.10)$$

$$\mu \ddot{\psi}_i(r,t) = -\frac{\delta E}{\delta \langle \psi_i |} + \sum_j \Lambda_{ij} | \psi_j \rangle \qquad (4.11)$$

where $-\frac{\delta E}{\delta \langle \psi_i |}$ are the forces that evolve the electronic system within an artificial Newtonian dynamics along with the nuclei. This allows us to avoid a costly self-consistent electronic ground state calculation at every time step. A critical assumption here is that the wavefunction that began in the ground state will remain there (or close enough) as the nuclei evolve in time. Another way of saying this is that the wavefunction remains close to the instantaneous ground state determined by the nuclear coordinates (i.e. the Born-Oppenheimer surface).

In order to ensure the energy-scale separation of the nuclear and electronic degrees of freedom (DOF), and to prevent energy transfer between them, the highest ionic phonon frequency w_I has to be much smaller than the lowest electronic analog w_e. This has been found proportional to:

$$w_e \propto \sqrt{\frac{\Delta E_{gap}}{\mu}} \tag{4.12}$$

where ΔE_{gap} is the KS single-particle energy gap [59]. Hence, the maximum integration time step depends on the fictitious electronic inertia like $\sqrt{\mu}$. Accuracy of the forces is the other constraint on the choice of μ. We consider the electronic subsystem to be most accurate when it is close to its instantaneous minimum energy configuration (i.e. the BO surface). A similar relation between μ and this deviation can be developed:

$$|\psi_\mu(r,t) - \psi_0(r,t)| \leq C\sqrt{\mu} \tag{4.13}$$

where $\psi_\mu(r,t)$ is the CP wavefunction as propagated by Eq.(4.11) and the electronic ground state is $\psi_0(r,t)$. Thus, μ acts as a way to tradeoff size of the integration timestep for accuracy to the instantaneous BO surface.

As already stated, μ has to be chosen small enough that $w_I << w_e$, where the high-frequency oscillations of Eq.(4.11) vanish on ionic time scales (i.e. $\langle \ddot{\psi}_i \rangle \simeq 0$). This ensures that the derivative of the instantaneous $\langle \Psi_\mu | \hat{H}_e | \Psi_\mu \rangle$, rather than fully-

minimized expectation value $\langle\Psi_0|\hat{H}_e|\Psi_0\rangle$ of the electronic hamiltonian yields the correct consistent forces. Thus, it is unnecessary to fully minimize $E[\{\psi_i\}; R]$, but rather to simply evaluate it for any time step. In contrast to BOMD, the computational cost to compute the nuclear forces is greatly reduced as there is no longer a need for a self-consistent loop to determine the minimum electronic ground state. However, the downside is that the time step must be small enough to ensure electronic and nuclear energy separation, which is typically much shorter than for BOMD.

As an added complication, if metallic systems are treated, w_e via Eq.(4.12) is zero so the electronic modes of oscillation necessarily overlap with the nuclear modes resulting in energy transfer between the two. A thermostat for the electronic degrees of freedom must be used to counteract the exchange of energy with the nuclear DOF [60].

Nowadays, it is more common to use BOMD in general. Theoretical advances have made wavefunction predictors very good, which keeps the cost of finding the ground state at every timestep relatively low.

4.3 Thermodynamic integration

To determine the free energy of Ta over a 2D phase space of volume and temperature F(V,T) we employ the method of thermodynamic integration. We cannot get free energies directly from MD, however, we can get free energies from exact analytical limiting cases, e.g. ideal gas or harmonic solid. We can then integrate the free energy changes w.r.t. temperature or volume via two thermodynamic relationships [61]:

$$P = -\left(\frac{\partial F}{\partial V}\right)_T \tag{4.14}$$

$$E = \left(\frac{\partial (F/T)}{\partial (1/T)}\right)_V \tag{4.15}$$

where P is the pressure and E is the total energy (kinetic and potential) of the system, both of which are accessible from MD as functions of volume and temperature.

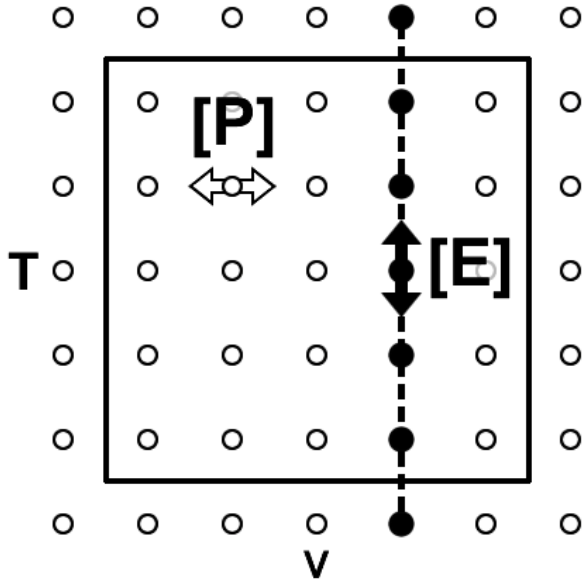
As a general procedure, we calculate pressure on a 2D grid of points and the total energy on a line at a constant volume in (V,T) phase space as illustrated in Fig.(4.1). The grid of points should be dense enough so the data fits $P_f(V,T)$ and $E_f(T)$ are smooth analytic functions. For the $E_f(T)$ curve, Eq. (4.15) is of such a form that, if satisfied by some $F_f(T)$, it is also satisfied by $F_f(T) - TS_o$ for some arbitrary entropy constant S_o. If this constant is known then the free energy F can then be integrated anywhere in the (V,T) region of interest. Furthermore, if we have a good model for P(V,T) we can fit its coefficients to *ab initio* data and extrapolate the exact free energy from an analytical model into the regions of interest. For example, for the fluid phases, we can integrate the free energy from an ideal gas starting point or potentially from a statistical mechanics model such as the Birch-Murnaghan or Peng-Robinson EOS.

As discussed in detail later, S_o can be determined by knowledge of the solid-liquid melting point at atmospheric pressure T_m. This temperature can be calculated experimentally or in a variety of fully *ab initio* ways [46, 49, 50, 61–64]. In the case of Ta, this problem was already studied with success, so we take this reference from previous *ab initio*-based calculations.

4.3.1 Finite temperature DFT

Finite-temperate DFT is important to this work because we are concerned with high temperature behaviour and at these temperatures the effects of electronic excitation in terms of delocalization and electronic entropy are non-negligible. In MD the temperature of the system is a measure of the kinetic energy of ionic motion (T_i). At zero Kelvin the total energy of the system is variational and involves an integral over the filled parts of the bands:

Figure 4.1: Thermodynamic integration of the free energy in a egion of two-dimensional (V,T) space. Pressure is determined on a 2D grid of points for integration in the left-right direction (white dots). The total energy is generated on an overlapping line of points for integration in the up-down direction.



$$E = \sum_n \frac{1}{\Omega_{BZ}} \int_{\Omega_{BZ}} \epsilon_{nk} \Theta(\epsilon_{nk} - \mu) dk \tag{4.16}$$

where $\Theta(x)$ is the Dirac step function. In finite-temperature DFT, electronic excitation effects are included as fermi-type smearing of occupations of single particle levels around the fermi level by replacing $\Theta(x)$ by:

$$f(\frac{\epsilon - \mu}{\sigma}) = \frac{1}{exp(\frac{\epsilon - \mu}{\sigma}) + 1} \tag{4.17}$$

where $f(\frac{\epsilon - \mu}{\sigma})$ is the occupation of an electron wavefunction of energy ϵ within a system with a fermi-level μ. In the context of Fermi-Dirac statistics, the amount of smearing (δ) can be interpreted as a measure of the electronic temperature T_e:

$$\sigma = k_B T_e \tag{4.18}$$

which is set equal to the desired ionic temperature T_i.

In this case though, the total energy is no longer variational. It is necessary to replace the total energy of Eq.(4.16) by a generalized free energy:

$$F = E - \sum_n \frac{1}{\Omega_{BZ}} \int_{\Omega_{BZ}} \sigma S_e(f_{nk}) dk \tag{4.19}$$

where S_e is the electronic entropy given by:

$$S_e(f_{nk}) = -k_B \sum_{nk} f_{nk} ln(f_{nk}) + (1 - f_{nk}) ln(1 - f_{nk}) \tag{4.20}$$

the calculated forces are now derivatives of the free energy F.

In practice, initially the ions are randomly distributed in a given volume and given a Maxwell-Boltzmann distribution of velocities corresponding to a desired temperature. Since at this point the potential energy of the system is rather poorly correlated to the kinetic energy (their "temperatures" do not correspond as the equipartition theorem tells us they should), the system must be thermalized. In this project we are employing a NVE rather than NVT ensemble so there is no thermostat available. Hence, thermalizing involves performing a MD run over time where there must necessarily be a rebalancing of kinetic and potential energy. The run is continued until they both stop fluctuating within a small tolerance. Now we measure the kinetic energy of the ions to get the effective T_i and set the amount of smearing δ so that T_e corresponds to the same temperature.

4.3.2 Statistical modelling of equation of state

For exploration of such a large region of phase-space, it would be very useful to be able to generate data through use of statistical modelling. This is both from a computational efficiency point of view as well as a means of extending the EOS to regions not easily accessed via DFT (such as very low density regions). In using TI in this project, we need to calculate the pressure for a wide range of volumes and

temperatures. Ideally, if we have a good model for the pressure P(V,T) we can fit its coefficients to *ab inito* data and integrate the free energy over a (hopefully) large region where the model is valid.

The tentative choice for this task was the Peng-Robinson EOS (PR EOS) which is a cubic EOS similar to the van der Waals model with the attractive term modified [65]:

$$\begin{aligned} P(V,T) &= \frac{kT}{v-b} - \frac{a\alpha}{v(v+b)+b(v-b)} \\ a &= 0.45724\frac{(kT_c)^2}{P_c} \\ b &= 0.07780\frac{kT_c}{P_c} \\ \alpha &= (1+0.37464(1-(T/T_c)^{\frac{1}{2}}))^2 \end{aligned} \tag{4.21}$$

where b is related to the size of the atoms in the repulsive pressure term and the product $a\alpha$ represents a measure of the interatomic attraction with the temperature dependence incorporated into α. Cubic EOS are advantageous for their simplicity of both application as well as fitting. Here we only need two free parameters to fit the model: T_c and P_c. Additionally, for this project it would be helpful to have a model for the low density fluid region of the EOS (due to DFT inaccuracies with low-density regimes in general). The PR EOS is typically chosen when accurate densities in non-polar fluids within the vapor-liquid equilibrium region are required and is a good balance between complexity and precision.

From this model we can derive an analytical expression for the free energy. For a non-ideal gas the Gibbs free energy can be expressed as a function of the fugacity f, reference energy G^o and pressure p^o:

$$\begin{aligned} G &= G^o(T,p^o) - kT\ln p^o + kT[\ln p + \ln(\frac{f}{f^o})] \\ &= (G^o(T,p^o) - kT\ln p^o) + kT[\ln p + \ln(\frac{f}{p})] \end{aligned} \tag{4.22}$$

the first group of terms can be determined from the statistics of an ideal monotomic gas:

$$\begin{aligned} G^o(T,p^o) &= E_0 - kT\ln(\frac{kT^{C_p}}{p^o\Phi}) \\ C_p &= 5/2 \\ \Phi &= T^{3/2}\Lambda^3 \\ \Lambda &= \frac{h}{\sqrt{2\pi mkT}} \end{aligned} \tag{4.23}$$

where Λ is the thermal de Broglie wavelength. We still need an expression for the fugacity which is provided by the PR EOS:

$$\begin{aligned} \ln\frac{f}{p} &= Z - 1 - \ln(Z-B) - \frac{A}{2\sqrt{2}B}\ln\left(\frac{Z+2.414B}{Z-0.414B}\right) \\ A &= \frac{a\alpha(T/T_c)p}{(kT)^2} \\ B &= \frac{pb}{kT} \\ Z &= \frac{pv}{kT} \end{aligned} \tag{4.24}$$

returning to Eq.(4.22), we can now express the Gibbs free energy for a non-ideal gas within the PR EOS as:

$$
\begin{aligned}
G = E_0 + kT(-\ln(C_1 T^{3/2}(v-b)) - 1 - C_2\alpha\frac{T_c}{T}\ln(\frac{v+2.414b}{v-0.414b}) \\
C_1 = \frac{(2\pi mk)^{3/2}}{h^3} = 4.5744 \times 10^{29}(m^3K^{3/2})^{-1} \\
C_2 = \frac{1}{2\sqrt{2}}(\frac{0.45724}{0.07780}) \\
b = 0.07780\frac{kT_c}{p_c}
\end{aligned}
\tag{4.25}
$$

Once we have fit the PR model with T_c and P_c we can determine the Gibbs free energy except for the free parameter E_0. We can establish this by equating the free energy of the solid and fluid phase at the melting point. Due to the thermodynamic integration procedure we are using (needing constant volume simulation cells), in this work we are actually dealing with the Helmholtz free energy rather than Gibbs. However, since $A = G - pv$, as the pressure goes to zero $A = G$ and we can still determine E_0.

4.4 Solid phase modelling

For this project the solid phase is modelled as a quasiharmonic solid which is both inexpensive compared to molecular dynamics and accurate below the Debye temperature (the inaccuracies above the Debye temperature is addressed in section 4.6.2). Under the quasiharmonic approximation (QA), the free energy per particle can be expressed as:

$$F_{QA}(V,T) = F_0 + \frac{E_0(V)}{N} + F_e(V,T) + F_v(V,T) \tag{4.26}$$

where $E_0(V)$ is the total energy at T=0 as determined from DFT. F_0 is an arbitrary offset to the free energy and N is the number of atoms in the simulation cell. F_e is the contribution of the electronic free energy given by:

$$F_e(V,T) = \frac{1}{N}(E_e(V,T) - TS_e(V,T)) \tag{4.27}$$

which includes the electronic entropy of Eq.(4.20) and E_e which represents the potential energy needed to thermally excite the electrons:

$$E_e(V,T) = \int [f_V^T(\epsilon) - f_V^0(\epsilon)]\epsilon g_v(\epsilon) d\epsilon \tag{4.28}$$

here $f_V^T(\epsilon)$ is the (V,T)-dependant Fermi-Dirac occupational density of the single-particle states and $g_v(\epsilon)$ is the volume-dependant density of states.

Finally, F_v is the contribution of vibrational free energy:

$$F_v(V,T) = \frac{k_B T}{N} \sum_m ln \left[2sinh \left(\frac{h v_m(V)}{2k_B T} \right) \right] \tag{4.29}$$

where we have a sum over stable phonon frequencies $\{v_m\}$ as calculated from a perturbative method, the specifics of which will be discussed next.

4.4.1 Phonon calculation

Calculating the phonon modes and occupations are crucial in using the QA in determining the free energy of the solid phase. As a first step, the total potential energy of the nuclei V_N is considered to be a function of the positions of all the ions in the crystal $R = X + U(t)$ where X is the atomic equilibrium position and U is the fluctuation around this position. If we expand the potential energy as a Taylor series with respect to atomic displacements U:

$$V_N = V_0 + \sum_{m\mu} \sum_i \Phi_i(m\mu) U_i(m\mu) + \frac{1}{2} \sum_{m\mu m\nu} \sum_{ij} \Phi_{ij}(m\mu, n\nu) U_i(m\mu) U_j(n\nu) + ... \tag{4.30}$$

where Φ are the derivatives of the potential energy w.r.t. the ionic displacements (i.e. force constants), i,j are the Cartesian components, m,n are the unit cell indices, and μ, ν enumerate the atoms within the unit cells.

The force constants are simply the derivative of the potential energy w.r.t. to the ionic displacements from the equilibrium positions:

$$\Phi_i(m\mu) = \left(\frac{\partial V_N}{\partial U_i(m\mu)}\right) \tag{4.31}$$

$$\Phi_{ij}(m\mu, n\nu) = \left(\frac{\partial^2 V_N}{\partial U_i(m\mu) U_j(n\nu)}\right) \tag{4.32}$$

since at equilibrium all the forces in a crystal must vanish, the first-order first constants of Eq.(4.31) must vanish. Furthermore, in the harmonic approximation we ignore the terms of order three or greater in the ionic displacements, so we are left with only the second order terms in the expansion. The total hamiltonian becomes:

$$H_{har} = T + V_0 + \frac{1}{2}\sum_{m\mu n\nu}\sum_{ij}\Phi_{ij}(m\mu, n\nu)U_i(m\mu)U_j(n\nu) \tag{4.33}$$

where μ, ν are the internal atomic indexes. From this we can derive EOM for all the atoms in the system:

$$M_\mu \ddot{U}_i(m\mu)^2 = -\sum_{n\nu j}\Phi_{ij}(m\mu, n\nu)U_j(n\nu) \tag{4.34}$$

with M_μ the atomic mass, giving rise to 3N differential equations (with N the number of atoms in system). Fortunately, for periodic structures we can utilize a suitable ansatz to decouple these equations. This involves expressing the displacements $U_i(M)$ as a plane wave.

$$U_i(m\mu) = \frac{1}{\sqrt{M_\mu}}U_i(\mu q)e^{i(q\cdot R_m - wt)} \tag{4.35}$$

where R_m is a lattice point. By considering solutions with a plane-wave form in the EOM of Eq.(4.34) we arrive at:

$$-w^2 U_i(\mu q) = -\sum_{\nu j}\sum_{n} \frac{1}{\sqrt{M_\mu M_\nu}} \Phi_{ij}(m\mu, n\nu) e^{iq\cdot[R_n - R_m]} U_j(\nu q) \tag{4.36}$$

Due to translational invariance, the terms in the sum only depend on the difference $m - n$. After summation over n, the m dependence disappears and we can rewrite this equation as:

$$-w^2 U_i(\mu q) = -\sum_{\nu j} D_{ij}(\mu\nu) U_j(\nu q) \tag{4.37}$$

where $D_{ij}(\mu\nu)$ are the elements of the so-called dynamical matrix (the Fourier transform of the force constant tensor). We are left with a linear homogeneous set of equations of order 3α where α is the number of atoms in a unit cell. For each q we can find 3α solutions for $w(q)$ at each q, and thus the phonon modes can be mapped.

To determine the phonon modes in our system we just need to determine the force constants which define the dynamical matrix. The Hellmann-Feynman theorem gives us a means to do this within a quantum mechanical description of a system. If the energy E of a wavefunction Ψ depends on a parameter λ then the derivative of the energy is related to the expectation value of the derivative of the hamiltonian as:

$$\frac{\partial E}{\partial \lambda} = \langle \Psi(\lambda) | \frac{\partial \hat{H}}{\partial \lambda} | \Psi(\lambda) \rangle \tag{4.38}$$

Thus, the force on nucleus I at R_I is:

$$F_I = -\frac{\partial E}{\partial R_I} = -\langle \Psi | \frac{\partial \hat{H}}{\partial R_I} | \Psi \rangle \tag{4.39}$$

for which Feynman showed to be dependent only on the electronic density n(r) and nuclear coordinates:

$$F_I = -Z_I \left(\int dr n(r) \frac{r - R_I}{|r - R_i|^3} - \sum_{J \neq I} Z_J \frac{R_I - R_J}{|R_I - R_J|^3} \right) \tag{4.40}$$

where Z_I is the nuclear charge.

So the general procedure is to displace each ion in our system by a small amount along each Cartesian axis. Then we can calculate the ground state electron density for this displaced-ion system with DFT. With the density we can calculate the new force on the displaced nucleus and subsequently the force constants. Finally, with the force constants we can determine the dynamical matrix and diagonalize to find the phonon spectrum.

4.5 Transport properties via Green-Kubo relations

As a secondary objective of this project we are concerned with determining the shear viscosity of Ta. In establishing the EOS we calculate small-scale thermal motion snapshots of Ta atoms. What is needed is a way to link this data to viscosity which is a bulk property of the system. That linkage is provided by the fluctuation-dissipation theorem of statistical mechanics.

Even systems in equilibrium, at non-zero temperature, still experience random fluctuations. The fluctuation-dissipation theorem states that the dissipation of these fluctuations has the same origin as the relaxation towards equilibrium of the system after it has been disturbed by an external force. Thus, both the dissipation and relaxation time scales are linked to the same transport coefficients. Green and Kubo showed that transport coefficients like diffusivity and viscosity are related to the correlation function of the corresponding flux [53,66]. In this way, the Green-Kubo (GK) relations link the macroscopic transport properties to microscopic fluctuations of the equilibrium distribution.

GK relations can be made for any transport coefficient in terms of its conjugate flux. We shall derive the GK relation for diffusivity as it is the most straightforward. Fick's law states that flux J is proportional to the concentration gradient:

$$J = -D\nabla c \tag{4.41}$$

where D is the diffusivity. From the continuity equation, we also know the time derivative of the concentration plus the divergence of the flux needs to be constant:

$$\frac{\partial c}{\partial t} + \nabla \cdot J = K \tag{4.42}$$

Setting the constant K conveniently to zero and substituting Fick's law we arrive at a PDE for the concentration profile:

$$\frac{\partial c(r,t)}{\partial t} = D\nabla^2 c(r,t) \tag{4.43}$$

This is our starting macroscopic relationship. If we now multiply both sides of Eq.(4.43) by r^2 and integrate over all space:

$$\frac{\partial}{\partial t}\int d\vec{r} r^2 c(r,t) = D\int d\vec{r} r^2 \nabla^2 c(r,t) \tag{4.44}$$

Through integration by parts and assuming our concentration function is normalized, in 3 dimensions the integral on the right is $6D$. On the left hand side, inside the integrand we have r^2 multiplied by the concentration. If we consider the concentration to be equivalent to the probability of having a particle at position r at time t then it becomes clear that the integral is just the average value of r^2 over all space:

$$\frac{\partial}{\partial t}\langle r^2(t)\rangle = 6D \tag{4.45}$$

We now have a relation between the diffusion coefficient D and the mean square displacement of the particles (microscopic property). If we introduce an alternative definition of the displacement in terms of the velocity:

$$r(t) = \int_0^t dt' v(t') \tag{4.46}$$

we can rewrite the mean square average displacement as:

$$\begin{aligned}\langle r(t)^2\rangle = \left\langle \left(\int_0^t dt' v(t')\right)^2 \right\rangle &= \int_0^t dt' \int_0^t dt'' \langle v(t')v(t'')\rangle \\ &= 2\int_0^t dt' \int_0^{t'} dt'' \langle v(t')v(t'')\rangle\end{aligned} \tag{4.47}$$

so the spatially averaged mean square displacement becomes an integral of the average product of the velocity at one point in time and the velocity at a different instant. This is called the velocity autocorrelation function. Returning to Eq.(4.45), if we take the time derivative we lose one of the integrals:

$$\lim_{t\to\infty} \frac{\partial}{\partial t}\langle r^2(t)\rangle = \lim_{t'\to\infty} 2\int_0^{t'} dt'' \langle v(t')v(t'')\rangle = 6D \tag{4.48}$$

If we consider that the system is translationally invariant with time we can shift t'' to zero:

$$\langle v(t')v(t'')\rangle = \langle v(t'-t'')v(0)\rangle \tag{4.49}$$

we now have our final result relating diffusivity to the autocorrelation of the velocity which is itself a measure of the way the velocity fluctuations in the system become uncorrelated (or dissipate) with time:

$$3D = \lim_{t'\to\infty} \int_0^{t'} dt'' \langle v(t'-t'')v(0)\rangle = \int_0^{\infty} d\tau \langle v(\tau)v(0)\rangle \tag{4.50}$$

For small values of τ we expect the autocorrelation function to be close to 1 as there has not been enough time for the velocities to change from one instant to the next. As we progress in time to large values of τ we expect the correlation to disappear and the autocorrelation to drop to zero as it is the ensemble average of all the velocity products in the system. What happens for intermediate value of τ is related to the dynamics of the particular system.

Considering that the shear viscosity η can be related to the xy component of the

pressure tensor:

$$P_{xy} = -\eta \frac{\partial v_x}{\partial y} \tag{4.51}$$

which is similar to Eq.(4.41) with a transport coefficient η and conjugate flux P_{xy}. A similar, though mathematically more obtuse, derivation can be performed for the viscosity in terms of the stress tensor, and we arrive at an expression similar to Eq.(4.50):

$$\eta = \frac{V}{k_B T} \int_0^\infty dt \langle P_{xy}(t) P_{xy}(0) \rangle \tag{4.52}$$

In this project we produce time sequences of finite-temperature MD runs from which we can obtain the stress autocorrelation function, so we should be able to calculate the viscosity as a by-product of the EOS simulations (under the assumption that the length of the MD run is sufficient for the autocorrelation to drop to zero).

4.6 Multi-phase equation of state for Ta

We now have all the tools necessary to tackle the problem of generating a multi-phase equation of state for Ta. Here we choose to perform all the MD calculations within the BOMD framework. Since we have access to an efficient way to find the electronic ground state, the time-step advantage of BOMD over CPMD was a deciding factor here. As an additional consideration, BOMD is better suited to handling metallic systems as it is easier to maintain adiabatic separation between the electrons and nuclei in a 'zero-bandgap' system (see Eq. 4.12).

We apply a finite-temperature formulation of DFT within an NVE (micro-canonical) ensemble. Here the variational quantity to be minimized and kept constant is the sum of kinetic and potential energy of the system. A NVE ensemble is chosen because there is no simple way to vary the number of atoms N with *ab initio* methods, a fixed energy E is needed for transport calculations (via GK relations), and if the volume

V is varied then this will drastically increase the equilibration time.

4.6.1 Computational details

All the MD calculations are based on DFT as implemented in the VASP package [39]. PAW pseudopotentials [37, 38] were used, treating the $5p$ electrons as valence. The GGA-PBE functional was used to approximate exchange-correlation effects as this represents a good tradeoff between computational efficiency and accuracy. Transition metals as a group are known to pose special problems for DFT both in terms of obtaining a sufficiently fast and sufficiently accurate treatment. This is due in part to the problems with the typical DFT exchange-correlation functional in accurately describing degenerate states [67], which are common in these systems. However, for some transition metals, in particular Ta, the straightforward application of the generalized gradient approximation (GGA) has been shown capable of delivering high-accuracy results [46, 68]. Many solid properties, melting curve, and shock-wave experimental data are well reproduced, especially after the issue of diamond-anvil cell (DAC) measurement mismatch [69] is resolved. All this previous work provides assurance that the accurate sampling of thermodynamic properties of Ta at the atomistic scale and for a wide range of pressure and temperatures is given reliably via GGA. Indeed, our own calculations of the melting curve and liquid-gas critical point is in excellent agreement with experiment.

Plane-wave basis cutoff was set to 275 eV, resulting in a total energy error of no more than 7 meV per atom and an error in pressure of no more than 0.4% at high pressures. All runs were performed with a Monkhorst-Pack grid of reciprocal points on a 2x2x2 mesh except for the liquid region calculations where a single gamma-point sample proved sufficiently accurate.

The number of atoms in the simulation cell was 128 except in the dilute gas phase

region where we had to use less due to computational resource constraints. Simulations performed on a system of 250 atoms at the highest pressure and temperature regions, where the approximation in our setup would be most noticeable, produced no significant change in results.

The time steps for the MD runs were chosen to be 2 fs, except at very low pressures where we could increase it up to 6 fs without a significant drift in total energy. The length of the MD runs varied in the range of 1-20 ps, depending on the temperature and pressure of the simulation point.

Next we review the specific procedure used for each region of the EOS and present our results. The solid phase is primarily handled within the quasiharmonic approximation framework with corrections provided with TI. The fluid phase is handled entirely with TI with an extension into the dilute regimes with the virial EOS. We use the fact that the solid and liquid free energies must be equal at the zero-pressure melting point to fix the open constants in the liquid and solid expressions for the free energy.

4.6.2 Solid phase

It is well known experimentally that the only stable solid phase for Ta is BCC. As mentioned previously, the solid phase is modelled with the quasiharmonic approximation. Here it is important to note that the calculation of the phonon modes and spectrum are done with an electronic temperature effectively zero. As a result, a large portion of the error in QA resides in the lack of thermal excitation effects on the electrons. As the electrons become hotter they tend to smear out into less localized states, softening the bonding and affecting the phonon frequencies. There is also the issue that as the ions become hotter their potential becomes less harmonic, especially as the melting curve is approached. These effects can be quantified by calculating

the free energy via thermodynamic integration and comparing it to the predictions of QA (Eq. 4.26). As finite temperature DFT is used for the MD step, we would expect it to be accurate for high temperatures, though there is a divergence as we approach T = 0 due to the $1/T$ term in Eq.(4.15). The QA should be accurate for the low temperature regions as these are well approximated by a harmonic potential with the Bose-Einstein statistical distribution of phonon occupations.

As the free energy via the TI procedure is a computationally expensive process, it is only calculated using MD runs for two limiting volumes, 10.0 Å^3/atom where the pressure is in the Mbars range, and 19.5 Å^3/atom resulting in close to atmospheric pressure. These volumes were chosen so as to span the predicted solid phase region. To fix the open S_0 constants inherent in the TI free energy we match the free energy calculated from the quasi-harmonic approximation method of Eq.(4.26) to the TI values at these volumes and at temperatures of 1100K and 1800K respectively. These temperatures were chosen as a region where the two energy curves $F_{TI}(T)$ and $F_{QA}(T)$ have common tangency. To remove these residual errors of anharmonicity and phonon temperature dependence we add the corrections calculated at the two limiting volumes:

$$\Delta F(T) = F_{TI}(T) - F_{QA}(T) \tag{4.53}$$

and linearly interpolate them for the intermediate volumes between for temperatures approaching T_m.

As a sidenote, we found that as temperatures approach the melting point omitting the electronic contribution of Eq.(4.27) leads to a total error of 480 meV/atom and 130 meV/atom respectively. Including this contribution leads to improvement in the relative error that is largely volume independent, which also moves from a reduction of two thirds around 3000K to being completely eliminated at 9000K. At least for the case of Ta, QA is very accurate for practically all volumes and temperatures.

Despite the generally high accuracy of QA even at melting temperatures, one issue remains to be addressed. To cover a sufficient region of the solid phase space we need to perform zero-temperature based DFT calculations for a bcc lattice with the volume it would have due to thermal expansion at T_m. However, at T=0 such a volume is well beyond the sublimation point and the bcc solid is not stable. Therefore, the QA result cannot be fully trusted as some of the phonon modes may become unstable. In addition, since the reference point for merging the solid and liquid phases is the atmospheric pressure melting point (see Sec.4.6.3), we must know the precise volume at this point to minimize errors in the phase boundary. Therefore, we must ensure that the thermal expansion is in agreement with the MD results up to T_m.

As demonstrated in Figs. (4.2) and (4.3), expanding the volume beyond the T=0 equilibrium soon leads to increased uncertainties in the QA free energy. Although the error accumulates only up to 30 meV/atom close to T_m it artificially shifts the zero pressure volume by 0.5Å^3/atom relative to the volume as calculated from MD. These fluctuations disappear at low temperatures for all relevant volumes suggesting that the QA error is mainly due to entropic effects. This leads to a natural way to address this issue; add a small entropy-like term $-TS^*(V)$ to the free energy of Eq.(4.26) so that the equilibrium volume coincides with the one from MD at a single temperature point not far from T_m. Since thermal expansion is correct at low temperatures and the expansion line is smooth and monotonic, such a procedure should suffice. Fig. (4.2) shows the change in free energy after the correction, which shows excellent agreement between QA and TI. The lattice thermal expansion agreeing with the MD results for all the temperatures below T_m as shown in Fig. (4.3).

This form of correction also guarantees that the free energy difference between the QA and TI methods remains unaffected by volume. The TI procedure involves setting an arbitrary constant entropic term $-TS_0$ at a volume of interest. Therefore, if *a posteriori*, a S^* term is added to the QA free energy, then using $S_0 + S^*$ in solving

Figure 4.2: Correcting the QA free energy for large volumes. For volumes beyond 19 Å^3/atom F_{QA} (full dots) develop small fluctuations at high temperatures. We add a small entropy term at each volume in this range so the corrected free energy (open dots) recovers the MD-calculated thermal expansion point (arrow) at a chosen temperature (3000K, near T_m).

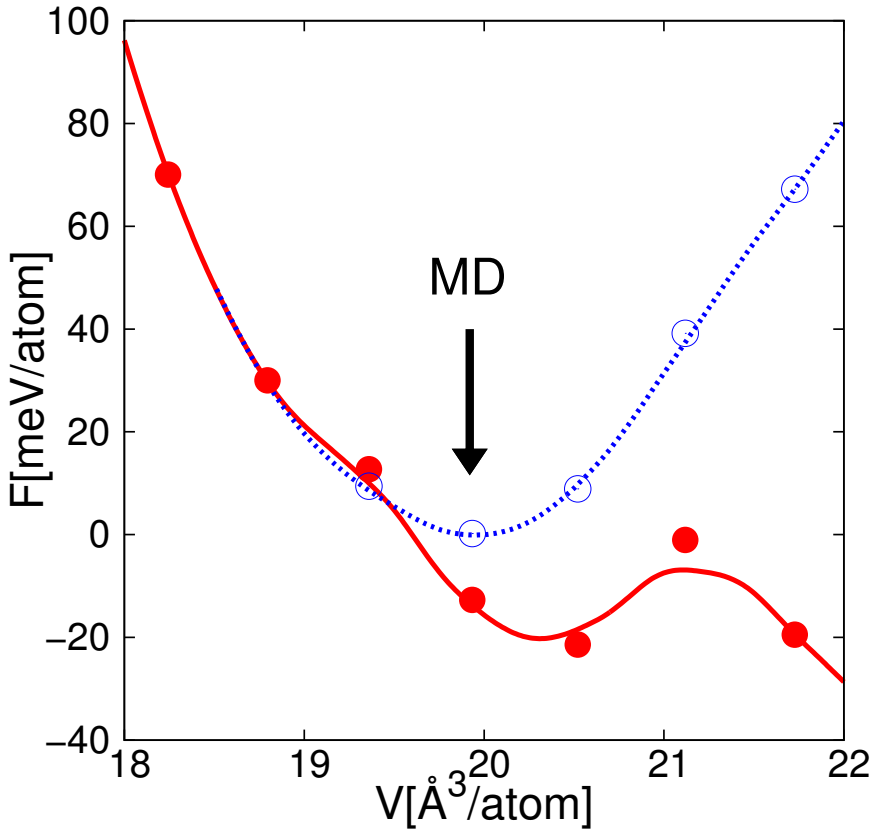


Figure 4.3: Correcting lattice thermal expansion of QA (red line) at atmospheric pressure and 3000K (dot) as described in the text. The corrected expansion (blue dotted line) recovers the MD result (green dashed line) for all temperatures.

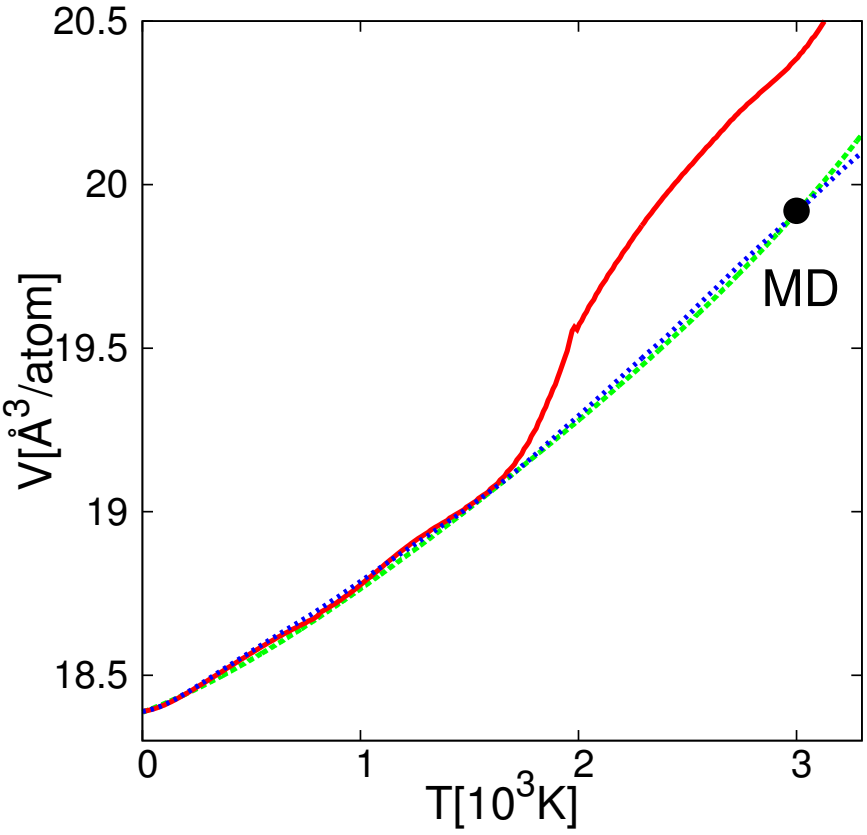
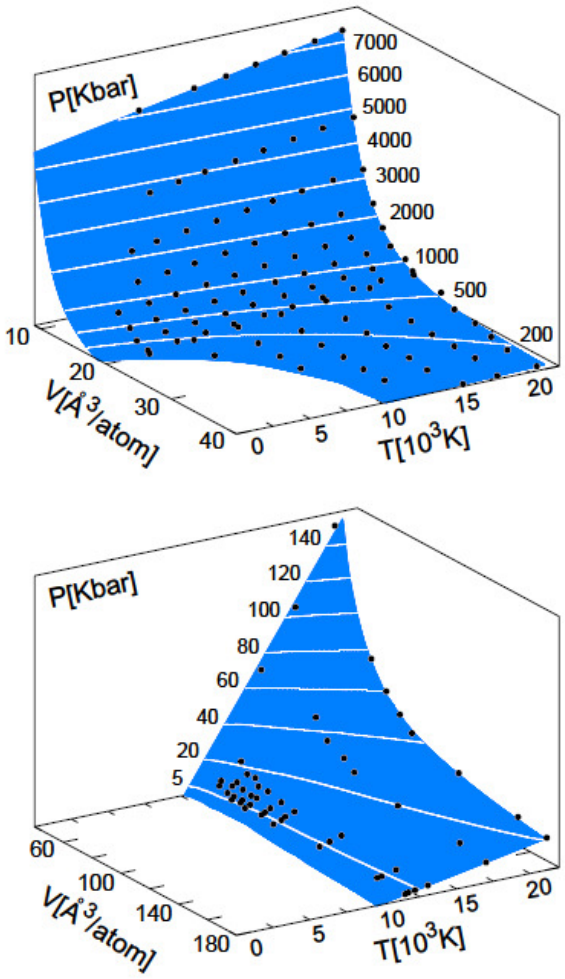


Figure 4.4: Fluid pressure field with DFT-generated pressure data points (dots). Isobaric lines are added as guides. Note changes of scale on the volume and pressure axes between the two figures.



Eq.(4.15) would keep the $\Delta F(T)$ of Eq.(4.53) difference unchanged. It is important to note that the two correction procedures are fully independent of each other, and can be employed separately.

4.6.3 Fluid phase

The free energy of the fluid phase is determined by TI as illustrated in Fig.(4.1). The pressure field P(V,T) is generated from *ab initio* MD runs at an array of volumes and temperatures - see Fig. (4.4). A significant challenge here is that the pressure field

varies over three orders of magnitude. In light of this, to achieve a smooth, analytic, and locally accurate fit for the pressure field we employ a fitting procedure as follows. We utilize a general 2D polynomial form for the pressure:

$$P(V,T) = \sum_{i,j}^{n,m} c_{i,j} V^i T^j \tag{4.54}$$

where the polynomial order and coefficients are best fit to the data. First we perform a global data fit, varying (n,m) to minimize the cross-validation score [70]. Then we uniformly divide the volume axis into 30 overlapping segments and perform a fit on each separately. To obtain good fits we add a number of points in-between the MD data points since their scattering is small enough to make such interpolation reliable. Finally, we smoothly connect the overlapping fits which now accurately represents all the actual pressure data.

Due to memory and runtime concerns, in order to reach the highest volume regions the number of atoms in the simulation cell had to be reduced to 64 above 40 Å^3/atom and 32 for volumes above 80 Å^3/atom. This is due to the inverse relationship between real and reciprocal space; as the real-space volume goes up, the reciprocal space volume goes down. Subsequently, for a given energy cutoff, there are many more plane-waves that must be used to represent the electron wavefunctions. To minimize the error involved with lowering the atom count we performed MD calculations at the maximum volume for each atom count used until the simulation became infeasibly slow. Our tests suggest that any error introduced in this fashion is negligible.

In theory, the free energy anywhere inside the single-component fluid phase can be determined entirely from TI from the ideal gas limit. This is due to the existence of a continuous path around the critical point. However, the solid phase free energy cannot be reached in this fashion. As a result, the solid-liquid free energy difference remains a free parameter. In principle, the solid-liquid free energy difference can be determined from the MD cohesive energy and TI alone integrating from analytic har-

monic and ideal gas models. However, such an approach is very sensitive to numerical and statistical noise due to the very long integration paths. To avoid this, we instead make use of the fact that the free energy of the solid and liquid phases must be equal at the melting point at atmospheric pressure T_m (here we assume atmospheric pressure is roughly zero pressure). At this point we directly match the free energies of the two phases and from there employ TI. In this way the integration paths are kept short improving accuracy.

The task of calculating T_m has been previously accomplished for Ta and we take the results of two particular publications [46, 71] that employ identical levels of *ab initio* theory to this work. The methodology used is a perturbative DFT correction to a classical force field (FF) result. Their predictions of 3326K and 3170K do not completely agree since they employ somewhat different starting force fields. Such a technique can be repeated several times until an average value arises. For our purposes, we took the median T_m=3248K as the established theoretical result. On the experimental side T_m varies in the range 3213-3290K [72, 73] so there is satisfactory agreement with our theoretical prediction.

At $T = T_m$ the integration procedure consists of the following. First, the free energy vs. volume curves along the melting isotherm for the solid and liquid phases (from the MD pressure field) are computed. From the requirement that both phases have the same free energy at zero pressure ($F_{solid} = F_{liquid}$), we offset the liquid curve so that the free energy matches the solid phase at the zero curvature point of the free energy curve (where $P = -\partial F/\partial V = 0$). The two volumes for which this occurs (one for the solid V_S, one for the liquid V_L) define the solid-liquid coexistence region. At this point we still need to determine the liquid free energy as we move away from the T_m isotherm. We select a volume V^* within the coexistence region (in this case 21.0 Å^3/atom) and integrate the liquid free energy with Eq.(4.15). An intermediate volume V^* is chosen so we are assured of being in the fluid phase regime in spite of small numerical errors in the exact location of V_L. We run a series of MD simulations

in the liquid phase and fit a polynomial to the E(T) datapoints. We then know the free energy of the liquid as a function of temperature at volume V^*. The complete 2D single-phase fluid free energy is then determined straightforwardly as an extension away from the V^* line at all temperatures, via the pressure field analytical integration of Eq.(4.14).

There is a problem that to extend the fluid phase into the ideal gas limit requires large simulation volumes with which DFT has difficulties. However, if we can reliably locate the critical point of our system (as discussed below), we can attempt to use a statistical model to extend our EOS. Initially, the hope was that a large section of the fluid phase could be expressed through fitting the PR EOS, and the solid and liquid free energies unified at the melting point as discussed above (Sec. 4.3.2). However, comparisons between the fitted and *ab initio* generated data was not favorable and the PR EOS was not used in this manner. Though the PR EOS turned out not to be an accurate enough model for liquid Ta, we can use a virial series [74,75] to extend the free energy into the ideal gas limit for all temperatures above the critical temperature T_c. The virial EOS has the form:

$$\frac{PV}{Nk_BT} = 1 + \sum_{i=1}^{n} a_i(T) \left(\frac{N}{V}\right)^i \tag{4.55}$$

For the fit we use the datapoints at and above the critical isotherm at 11600K as shown in Fig.(4.5). In order for the fit to closely follow the pressure at and just above the critical point, the fitting order n has to be set to 6, reflecting the relevance of many-body interations in this region. Here, however, signs of overfitting inevitably start to show due to the curvature of the critical isotherm being on the same order than the fluctuations of the fitted data. Thus, such a fit cannot be reliably used to extrapolate the virial EOS completely to the ideal gas limit, as further testing confirmed. When the fitting order was reduced to 4, extrapolation is very smooth and blends with the ideal gas limit perfectly. Thus the question of the best choice for n remains problem specific.

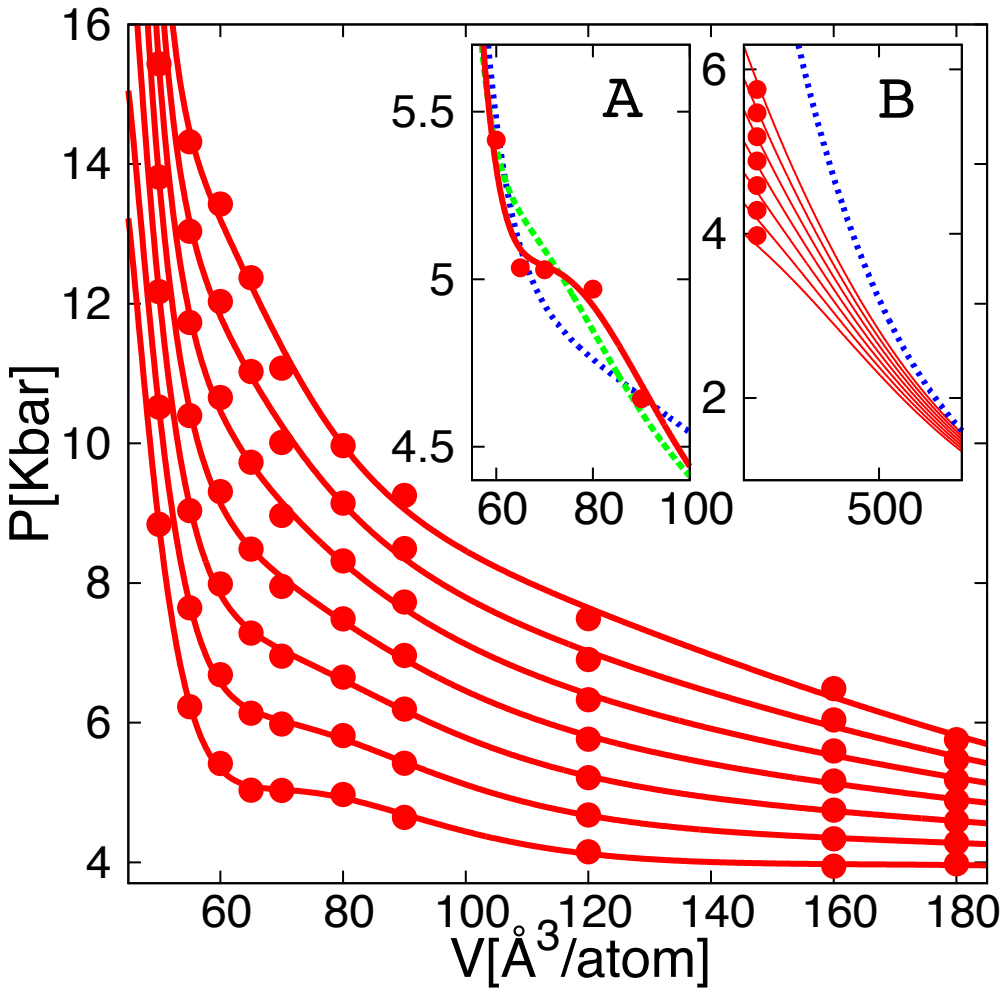
Below T_c we cannot use the virial EOS due to the phase boundary. However, are able to extract enough information to model a vapor pressure curve P_{vap}, extending the free energy calculation to include the liquid-gas two-component regime. This is discussed in more detail below.

In the remaining sections of this chapter we shall examine the phase coexistence regions of our EOS, specifically our prediction for the melting curve, critical point, and vapor pressure curve. This should further help us examine the accuracy and utility of our approach. For the most part, these are generated as by-products of the work in the previous two sections. What is striking is the excellent agreement with previous computational efforts in which these features were specifically sought for and which are not part of a unified EOS.

4.6.4 Melting curve

With solid and liquid phase data prepared with sufficient accuracy, developing the melting curve is a purely numerical procedure. This involves locating an array of common tangents between two sets of data across a range of temperatures. One complication is that the solid and liquid free energies are typically very close. Thus, the resolved locus of points are often more sensitive than the underlying data itself. At each temperature (as we evolve upward from T_m) we employ local quadratic fitting of the two sets of data. The common tangent is solved for analytically as connecting the two parabolas. Fig. (4.6) shows the result in the V-T plane as the two lines encompassing the solid-liquid coexistence region $V_S(T)$, $V_L(T)$. Since the fluid free energy is a fit over a collection of dispersed points of pressure data, both $V_{S,L}$ have oscillations. Running more MD points is too costly since the precise position of the melting curve is unknown beforehand and would require too many points to be run. We thus resort to *a posteriori* smoothing of $V_{S,L}$ by averaging them at each point over

Figure 4.5: Virial expansion of the pressure field. The pressure data (dots) is fitted (lines) starting from T_c in increments of 200K. Inset A: datapoints at the critical isotherm (dots) are fitted with polynomials of order 6 (red), 5 (green), and 4 (blue). Inset B: order 4 fits (full lines) can be reliably extrapolated toward much larger volumes. The highest T isotherm is shown meeting with its ideal gas counterpart (blue line). Volume axis is logarithmic.



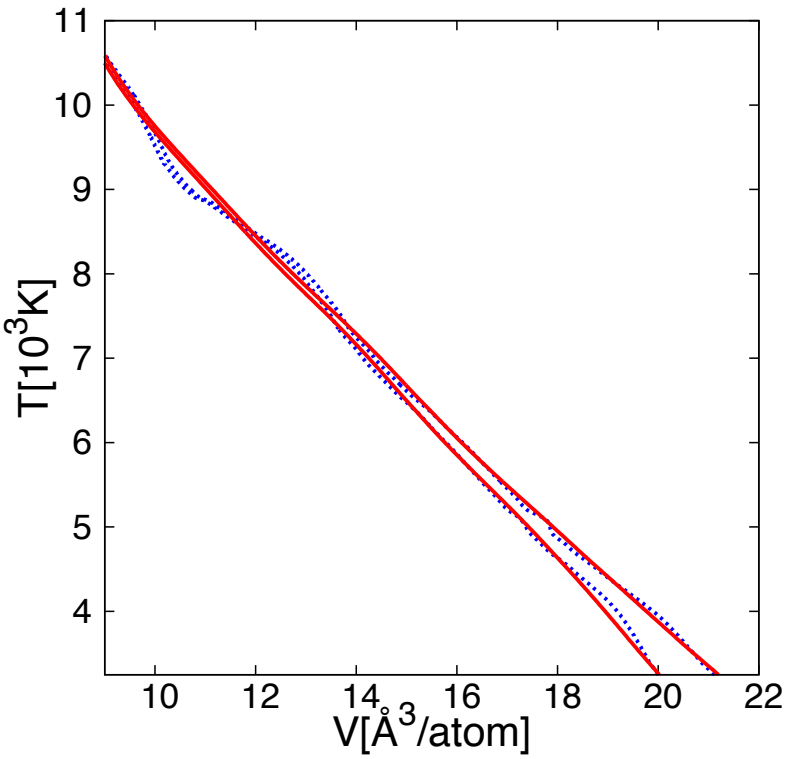
a surrounding interval. The minimum width of the interval is chosen so as to remove the local oscillations and still maintain the general trend of $V_{S,L}$. The melting curve is then fully resolved as a locus of points on the liquid pressure field at $P(V_L(T), T)$.

In the literature there are several publications calculating the melting curve of Ta, all done via classical atomistic simulations with *ab initio* input included on some level. These are shown in Fig. (4.7) as: (1) extended 19-parameter FF of embedded atom model (qEAM) type fitted mainly to zero-temperature DFT data, melting curve calculated from a coexistence method [49]; (2) model generalized pseudopotential theory (MGPT) with up to 4-body terms in a volume-dependent FF, melting calculated from direct melting within periodic cell (corrected for hysteresis effects) [50]; (3) reference EAM FF optimized at low and medium pressures in the liquid phase, with free energy perturbatively corrected for differences against the DFT hamiltonian, melting curve from coexistence simulations [46]. DFT input in all these approaches utilizes the same GGA functional as this work.

At zero pressure all these theoretical predictions fail within or close to the range of experiment (apart from the slight overestimate of MGPT). At high pressure only a single experimental point is available from shockwave compression measurements [76] at 3Mbar with a large error margin of 1500K. All simulation methods apart from MGPT agrees with this experimental point. However, MGPT does agree with qEAM and our work until pressures exceed 1.5 Mbar. Our results overlap qEAM data and extends them into higher pressures. Result (3) have a peculiar behavior of the dT/dP slope sharply decreasing for the mid-range pressures, though still in agreement with experiment at higher pressures.

The zero pressure dT/dP slope is a further consideration in comparing the various theoretical results with experiment. Experimentally, this is 6 ± 1K/Kbar [77]. Extracting such information from our curve requires additional degree of smoothing. We perform 2nd and 3rd order polynomial fits to the data shown on the right panel

Figure 4.6: Melting curve in V-T plane: V_s, V_L lines encompassing the solid-liquid coexistence region both before (dotted) and after (solid lines) the *a posteriori* smoothing procedure described in the text.

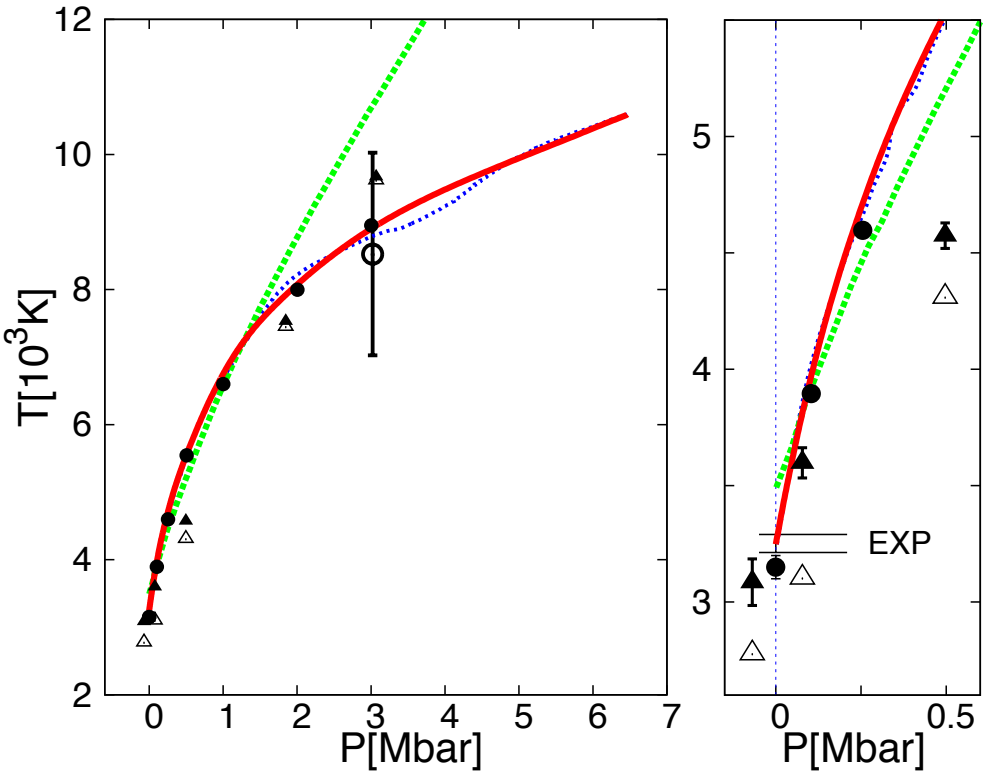


of Fig.(4.7) up to 0.6 Mbar. These give estimates of 6.6 and 7.7K/Kbar which is in good agreement with experiment.

4.6.5 Critical point

The critical point (CP) of the liquid-gas phase transition is an important, universal feature of an EOS, positioned between dense solid/liquid regions and the ideal gas limit. However, the CP lies at the frontier of the traditional domain of validity of DFT; as a method it is generally considered less reliable as the interatomic spacing increases towards the gas phase regime. As a further complication, experimental

Figure 4.7: Melting curve calculated from various approaches: (1) qEAM [49] (dots), (2) MGPT [50] (dashed line), (3) EAM (open triangles) with DFT corrections [46] (full triangles) and this work (red line after smoothing, blue before). At the high pressure enf, most theoretical predictions fall within the error bars of the single experimental datum (open circle). The right panel shows the atmospheric pressure region in more detail. Experimental T_m measurements are within the range of the two horizontal lines.



determination of the critical point is an open question for most metals [78] as it typically lies at extreme temperatures and pressures [79]. However, a number of previous experimental and semi-empirical estimates exist [44, 78, 80–85], though they differ considerably. As such, any clarification of the experimental results via our *ab initio* approach should prove particularly valuable. [79]

Direct calculations of the critical point of the liquid-gas transition in materials are traditionally done in the context of atomistic potential fitting, as they were previously considered too costly for a purely *ab initio* approach [86]. This is even more pronounced for strongly binding systems like Ta, where, as the temperature falls below T_c, the gas side of an isotherm soon moves to extremely large simulation volumes. Nevertheless, thanks to increases in available computational power and through the use of carefully set up MD runs, we can now perform this task via *ab initio* methods.

Through an initial sparse search of the pressure field across the low-density region of (V,T) phase space we can assess the rough location of the CP along the highest isotherm with $dP/dV > 0$. From there we calculate several surrounding isotherms in more detail with the following procedure. For each volume, starting just above the estimated T_c at the 13000K isotherm, we continue the MD simulation with the kinetic energy reduced via velocity rescaling to equilibrate at a new, lower ionic temperature target; the electronic temperature is reset to this value as well. Since the density is relatively low, the kinetic energy change will not significantly affect the potential energy, causing only a minor restructuring in the relative positions of the atoms. After a short thermalization process (few ps) we collect data for approximately 10-20 ps. By limiting the sampling time in this manner the system has a high probability of avoiding the onset of phase separation and remains in an undercooled fluid phase for the full duration of the data acquisition. Phase separation was nonetheless observed in a few cases as a sharp change in temperature and pressure - for these runs the whole procedure was repeated from a different starting configuration (i.e. from a later simulation time, but still from the 13000K isotherm). These results are collected

in Fig.(4.8). The vertical line divides the data calculated with a 64-atom cell on the left from the 32-atom cells on the right. To minimize error we have pushed this line as far to the right as we could before the simulations became prohibitive costly. We found that any bias due to the change in system size was small relative to statistical fluctuations in the data, allowing us to move seamlessly across data calculated with various simulation sizes.

We calculated six isotherms with temperatures in the range 10500K-13000K as shown in Fig.(4.8). As a consequence of the NVE ensemble, the final average ionic temperature in each case will deviate slightly from the desired target value. For that reason we fit the pressure data separately at each volume with a quadratic polynomial (which is demonstrated in the inset figure). These are the final P(V,T) representation of this part of the EOS. We can then use this two-dimensional pressure field to plot the intermediate isotherms in as fine a detail as needed to determine the CP as illustrated in Fig.(4.9). As before, the location of the CP is estimated by finding the highest isotherm with $dP/dV > 0$ at any one segment. The middle point of such a segment is then taken as (P_c, T_c, V_c). The error bars associated with the statistical fluctuations of the sampled pressure data are approximated by the following procedure. Assuming the fluctuations do not depend significantly on density and temperature in the vicinity of CP, we calculate the average standard deviation of the calculated points from their respective quadratic fit. In this study, this value is $\sigma^* = 0.058$Kbar. We then replace each calculated point by drawing from a normal distribution of σ^* width around each fitted line, and repeat the quadratic fitting procedure. A new CP location is estimated by the method above and we repeat this entire process 1000 times. The error bar estimates extend over the intervals which contain 2/3 of all the sampled CPs. Finding the error in volume is less straightforward because the volume sampling in our setup is performed at specific points. Here we can find the unique normal distribution which, if data were drawn from it, would produce the same discrete distribution of V_c samples. The mean and the width of such a distribution is then taken as the final V_c and the associated error. This, the fully *ab initio* CP parameters

Figure 4.8: *Ab initio* isotherms for critical point determination (full square). Direct MD averages (open circles) fitted by quadratics at every volume as shown in the insert (full circles). Thin vertical line divides 64 and 32 atoms/cell systems. The full black line shows the critical isotherm of a Peng-Robinson model with $\{P_c, T_c\}$ set to match the estimated *ab initio* values with its CP shifted to a higher volume (hollow square). The short thick line estimates a segment of the binodal line (see text). The volume axis is logarithmic.

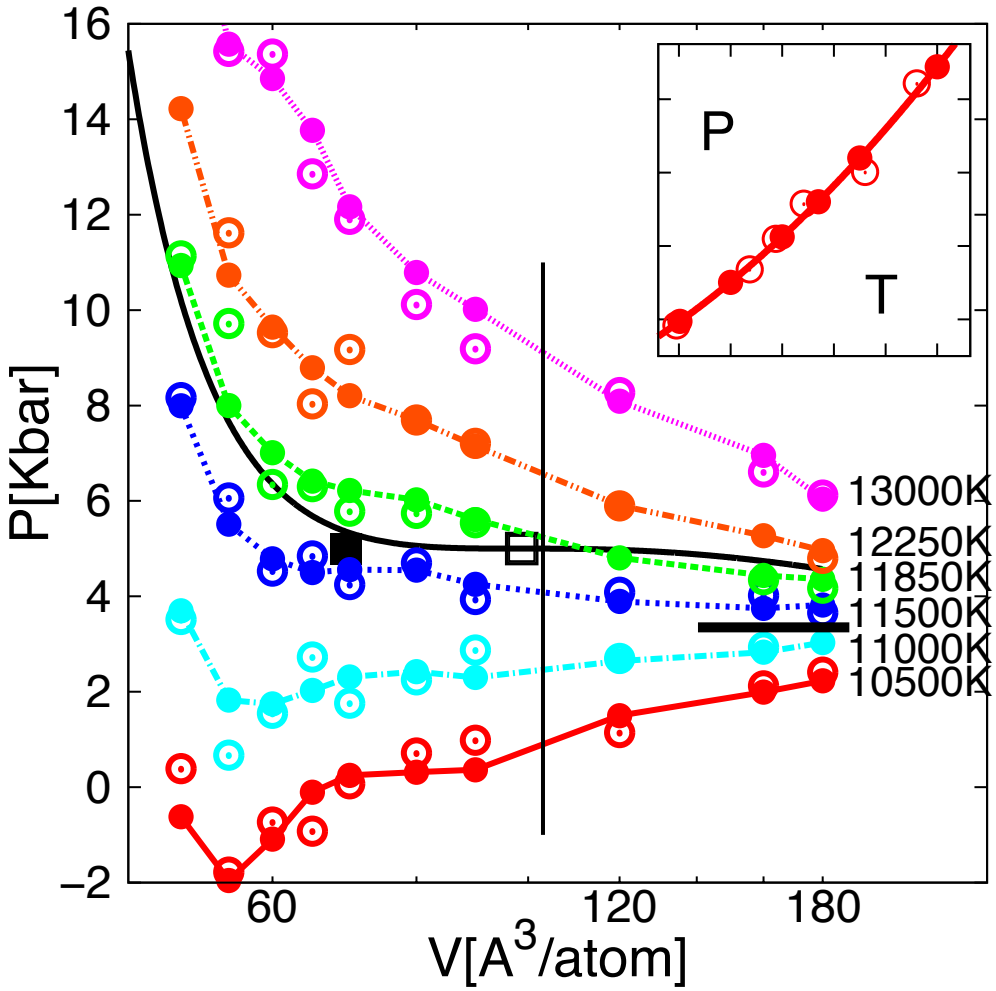
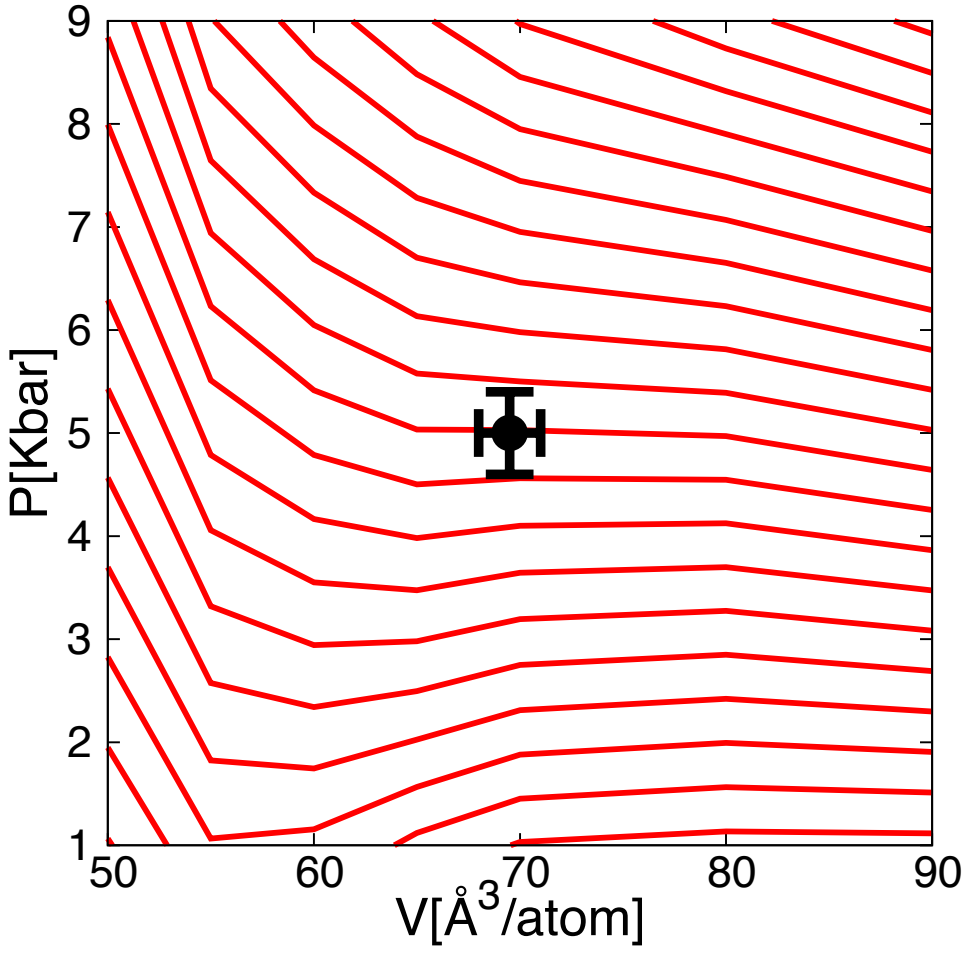


Figure 4.9: Array of fine detail quadratic pressure fits to determine the precise location of the CP (dot). The cross size reflects the statistical uncertainty in the predicted CP.



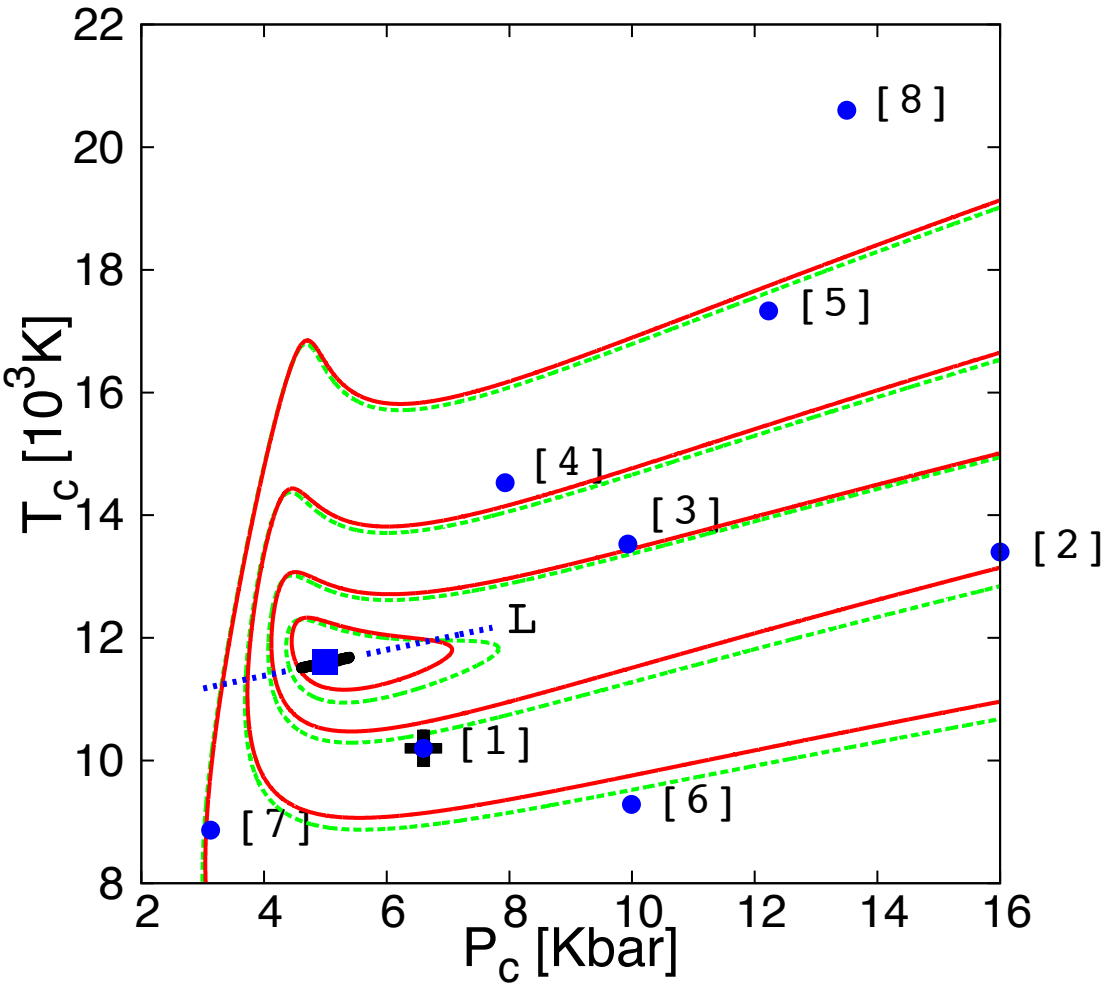
are estimated as $T_c = 11600 \pm 90K, P_c = 5.0 \pm 0.4Kbar, V_c = 69.5 \pm 1.5^3/atom$ with error bars shown in Figs. (4.9) and (4.10).

Accurate experimental determination of the CP of all but the simplest metals is an ongoing problem. For Ta, widely differing estimates have been offered over the years, both from theory and experiment, as can be seen in Fig.(4.10). Most of these estimates are indirect, semi-empirical approaches based on correlation of various metallic properties with the critical parameters, such as ionization potential, heat of vaporization, cohesive energy, etc. However, recently one experimental group performed the most direct measure so far of the high-rate heating of the metal [78], approaching the CP via the spinodal line, and can be expected to be of higher quality than previous attempts (point (1) in figure with error bars). For this reason it is satisfying that our *ab initio* estimate of CP is closest to this particular measurement.

4.6.5.1 Peng-Robinson EOS

In order to compare previous estimates of the CP, it is interesting to examine how the calculated pressure field in the vicinity of the CP compares with the prediction of a commonly referenced Van der Waals-type EOS. Here we choose the modified Peng-Robinson (PR) model [65] which typically performs well for density of non-polar liquids; this EOS is completely specified by two critical parameters $\{P_c, T_c\}$. The solid line in Fig.(4.8) is the critical isotherm of the PR model with the *ab initio* estimated values. The PR prediction of V_c is shifted to $98.7^3/atom$, and accordingly the pressure declines more slowly with volume. The PR model, although unable to precisely predict the pressure in the vicinity of the CP due to large discrepancy between its estimate of V_c and what we find with TI (see also the following section on vapor pressure), proves useful for studying the agreement with earlier estimates of P_c and T_c. Here we would like to go beyond the simple comparison of CP location and also consider the implied pressure fields. Simply put, two apparently distant CP's may be associated with very similar pressure fields P(V,T).

Figure 4.10: Various previous estimates of CP (dots) together with our *ab initio* estimate (square). Straight line L (dotted) passing through the estimated CP is the locus of statistically sampled (P_c, T_c) pairs as described in text (2/3 of all pairs fall within the thick segment of L, giving the error bars on the final CP estimate). Isolines of constant δP_{PR}^{DFT} (red) and δP_{PR}^{PR} (green) indicate pressure field simularity with the *ab initio* result.



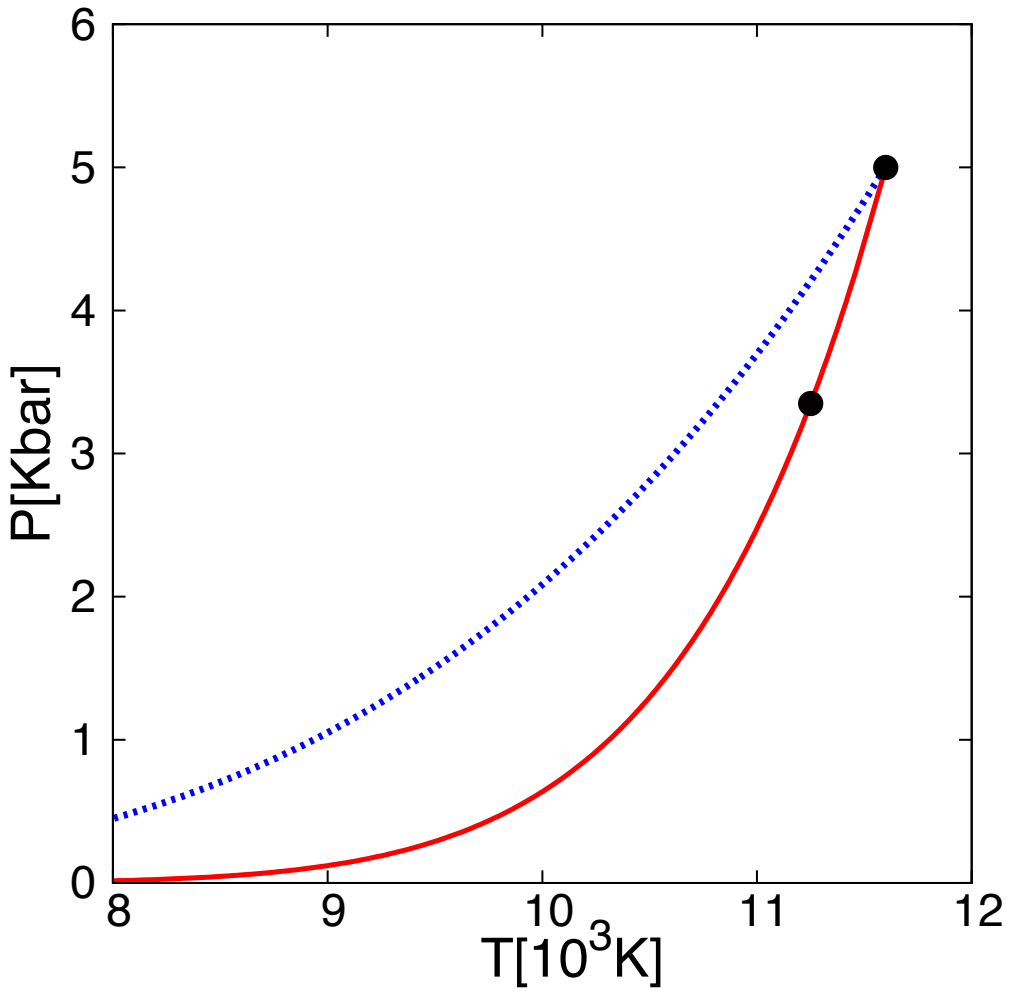
To this end, we calculate the standard deviation of the pressure difference between *ab initio* and PR pressures $\delta P_{PR}^{DFT} = \langle (P_{DFT} - P_{PR})^2 \rangle^{\frac{1}{2}}$ at a finite set of calculated datapoints (full circles in Fig. 4.8) above the critical isotherm as a function of PR parameters P_c and T_c. What we find is that δP_{PR}^{DFT} is minimized when the PR parameters and the *ab initio* calculated critical parameters $\{P_c, T_c\}$ closely coincide. This gives us a way to quantify the disagreement between two estimates of the CP in a metric proportional to the corresponding disagreement in physical observables (in this case, pressure). We can then use a previous CP estimate with the same PR model to generate a pressure field and assess its quality via deviation from *ab initio* or experimental pressures. This alternative metric is limited only by the precise prediction of the P_c, T_c values.

We can partition the P_c vs T_c diagram by iso-lines of sharply rising δP_{PR}^{DFT} (2, 4, 8, and 16Kbar) as a visual help in roughly dividing the previous estimates into three groups of descending agreement with the first principles result: (1-3) [44, 78, 82], (4-7) [80, 81, 83, 84], and (8) [85] - see Fig.(4.10). In this metric, result (2), although far in absolute P_c, T_c values, yields a very similar pressure field to the one generated via DFT, while exactly the opposite holds for estimate (7). Measurement (1) is closest in both criteria. In the previously described procedure of statistically sampling CP parameters to determine their error bars, individually sampled (P_c, T_c) pairs are highly correlated in the P-T diagram, with points closely falling on a straight line L, again as shown in Fig. (4.10). The direction of L very much agrees with the direction that minimizes δP_{PR}^{DFT} as PR parameters are varied. We thus interpret δP_{PR}^{DFT} as such: when testing the accuracy of different combinations of P_c, T_c close to L, the underlying pressure field resembles the one we have generated with DFT. Moving away from L yields a readily detectable change in the shape of the pressure field. Finally, we calculate, at the same set of points as before, δP_{PR}^{PR} using pressure differences between two PR models, a reference one fixed on *ab initio* parameters, and the other with its parameters varying. As Fig.(4.10) shows, this function does not significantly differ

from δP_{PR}^{DFT}, therefore for the analysis of the $P_c - T_c$ diagram the *ab initio* pressure field can be fully substituted with the PR one. This also implies that the gross behavior of the PR model is in good agreement with DFT around the CP; suggesting that a different cubic EOS may be able to generate a sufficiently accurate pressure field.

The two metrics we employ to assess the quality of previous CP estimates bring out interesting details. First, in both absolute value of the CP and in similarity of generated pressure fields, our data is closest to the only available direct measurement [1]. Confluence of these two independent results, a fully experimental and a fully *ab inito* one, strongly suggest that the true location of the CP can be greatly narrowed to their vicinity. Estimate [2] is based on cohesive energy argument and an empirical EOS composed from a large collection of experimental data. Although it overestimates P_c by more than twice, Fig. (4.10) shows that it implies a pressure field similar to ours, in turn suggesting a high general accuracy for the EOS. Result [3] also indicates a good quality pressure field but the authors provide few details regarding how this empirical estimate was made. The estimates from the next group in declining quality [4-7] all have the similar approach of employing atomistic models: [4] and [6] soft-sphere, [5] a hard-sphere Van der Walls, and [7] overlapping virtual atom models. All of which parametrized by indirect experimental data like liquid density near CP, cohesive energy, ionization potential, and scaling arguments. Interestingly, reducing the sophistication of these models does not introduce a significant decline in predicting CP, with soft-sphere models somewhat better performing. Finally, the semi-empirical estimate [8] based on heat of vapor data, overestimates both P_c and T_c by almost twice and implies a very inaccurate pressure field.

Figure 4.11: Vapor pressure curve as determined from *Ab initio* data (red line) as a fit using two datapoints (circles) lying on a binodal line. The upper point being the CP and the lower one being the estimate as described in the text. Also plotted is the Peng-Robinson model estimate of the vapor pressure curve (blue line) with the same $\{P_c, T_c\}$.



4.6.6 Vapor pressure curve

Below T_c we can extract enough information to model a vapor pressure curve, extending the free energy calculation to include the liquid-gas two component region. To estimate the vapor pressure in the liquid-gas coexistence region the standard approach is to fit the pressure using the Antoine equation [87]:

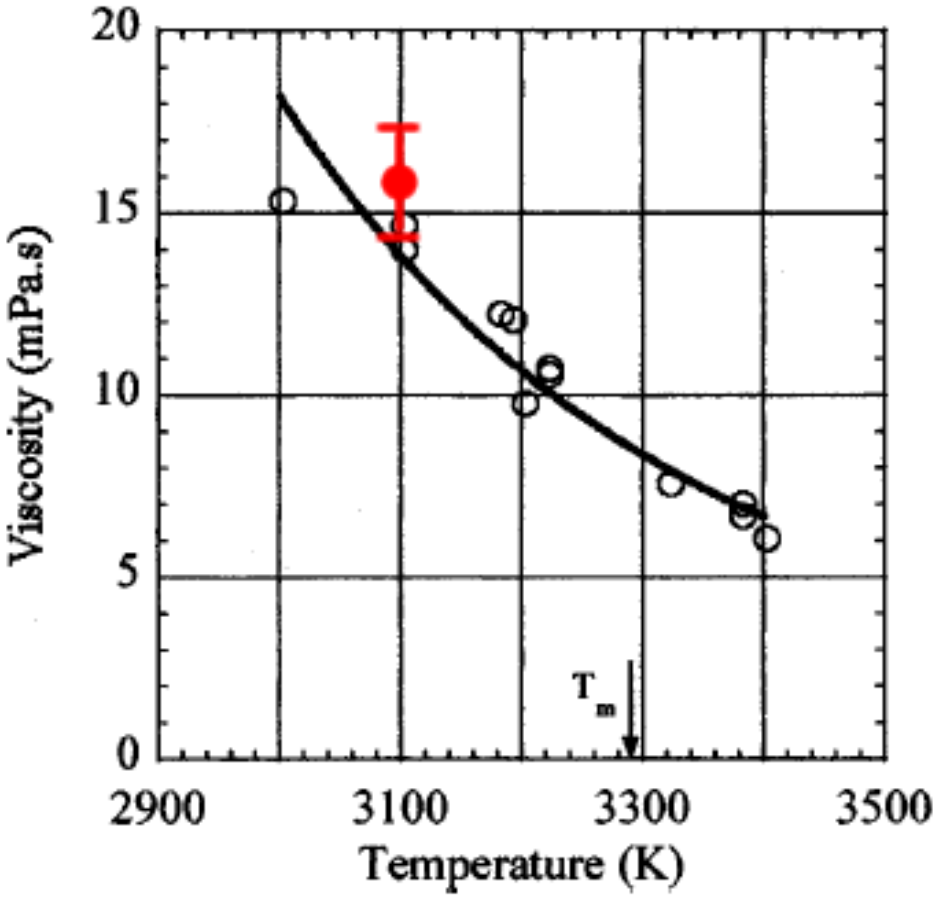
$$logP_{vap} = a - b/T \tag{4.56}$$

which is derived from the Clausius-Clapeyron relation [68]. Here we need two coexistence points to solve for the a,b parameters. One of these is the just the critical point. For the other we use the fact that the binodal line reaches very large volumes at pressures just below P_c. Since at the largest calculated volume the 11500K isotherm pressure decreases with volume (gas phase), while the 11000K isotherm pressure increases (coexistence phase), we can assume the binodal is trapped between these two lines at the largest calculated volume, giving the estimate $P_{vap}(11250K) = 3.35$Kbar. With this, the parameters are fitted by a=14.482 and b=149320K, the fit shown in Fig.(4.11). The binodal line to the left of the CP in the PV diagram is determined as the line where the liquid pressure equals P_{vap} across temperatures. These results are also compared to the PR model fit and we find that the *ab initio* binodal differs notably from the PR result. This is another reason the PR EOS was relegated to an evaluation tool with the virial EOS instead being used to extend the direct DFT results.

4.6.7 Transport properties

As developed earlier, the Green-Kubo equations are the mechanism through which we hope to determine the viscosity of Ta for various points in V-T phase space. This involves calculation the stress autocorrelation function (SACF) from MD runs at a specific temperature and volume. Typically a small number of long trajectories are used to generate the viscosity accurately. In our initial investigation, over the typical

Figure 4.12: Calculated viscosity from Green-Kubo relations. Point in red is the determined value from stress autocorrelation function of 100 individual trajectories, the open circles are determined via experiment.



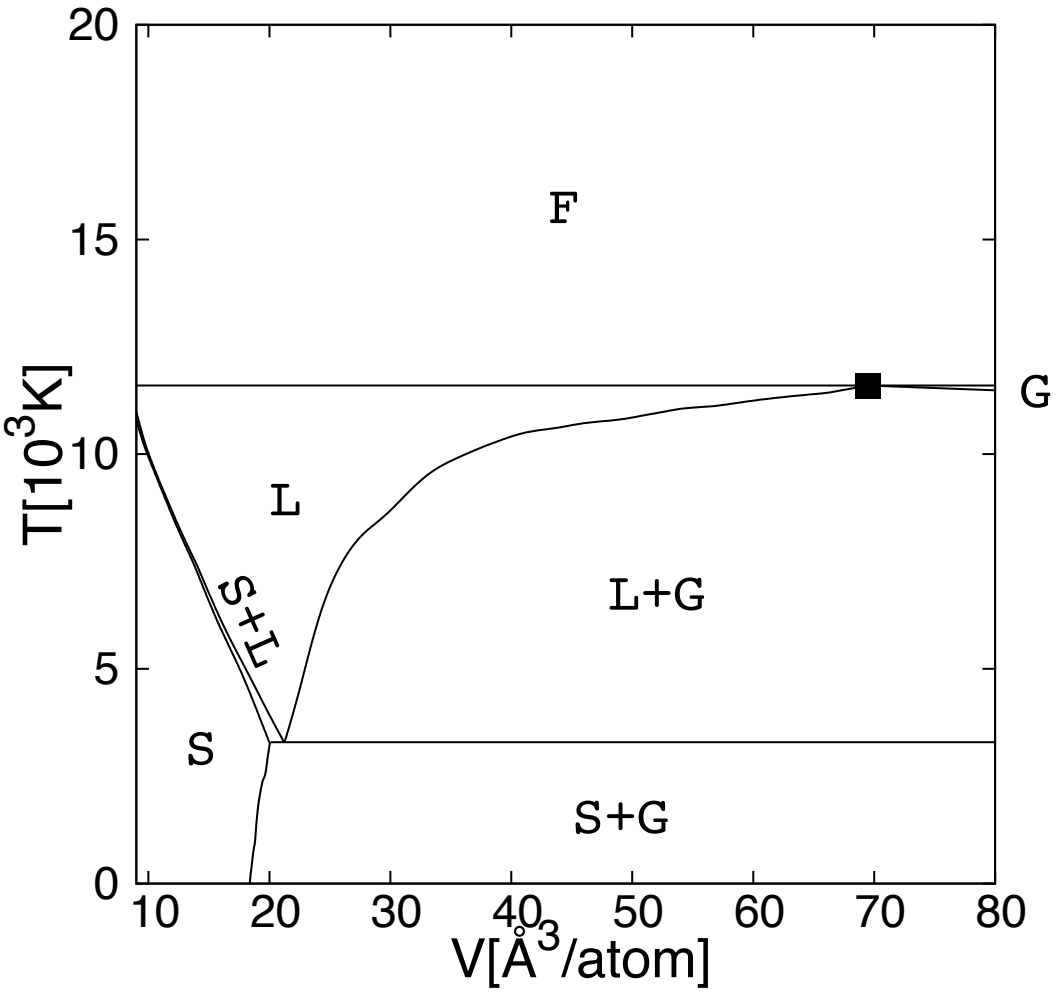
trajectory time-scale of our MD runs there is a lack of a long tail to the SACF. Waiting for the autocorrelation to fall to zero would require an unreasonably long MD run on the order of 100 picoseconds. A better option is to run many short trajectories (on the order of 100's of femptoseconds) which allow larger MD timesteps to be used as well as improve the statistical sampling of the system. This approach also lends itself to easy scaling as a single job can be applied to each computing node. This requires that each trajectory is properly thermalized to begin with, but this can be accomplished by preparing the starting configuration with a classical MD run which has the identical radial pair distribution function as our ab-inito liquid results.

Work has not been entirely completed as of the time of this write-up, but there are some promising initial results. We calculated the SACF's of 100 trajectories with a total runtime of 120 ps (MD timestep of 6fs). This data results in a calculated viscosity just slightly higher than recent experimental measurement [88] as can be seen in Fig. (4.12).

4.7 Conclusion

Compared to previous efforts, this is the first direct translation of the DFT Hamiltonian into thermodynamic equations of state over such a wide span of phase space. It can thus serve as a point of reference for calculations done with more approximate levels of quantum input. Analysis of the melting curve predictions is a natural place to start due to the amount of previous work and the relative ease in comparing results. Since all discussed theoretical calculations of the melting curve involve basically the same quantum input, all should give a similar output. This is especially true at lower temperatures where the qEAM and MGPT models are primarily fitted. The concurrence of our result with qEAM and MGPT below 1.5 Mbar strongly suggests that most DFT effects are fully captured at these pressures. The perturbative EAM+DFT result agrees with other methods in the vicinity of zero pressure, confirming the validity of this approach when the classical EAM part is not far off from the final result.

Figure 4.13: Complete *ab initio* phase diagram for Ta including solid (S), liquid (L), gas (G), and fluid (F) phases as well as their coexistence regions. The critical point is marked by a square dot.



However, near the 0.5 Mbar mark the EAM result underestimates the temperature by more than 1000K and at this point the quantum correction fails to return the full DFT result. Although more work would be needed to definitively investigate this issue, we can tentatively reason that at this pressure the classical portion of the model is not accurate enough to continue to apply the quantum effects as a perturbation.

Our result shows remarkable agreement with qEAM in all the reported datapoints. This is somewhat surprising since the FF was fitted to zero temperature DFT data, with the exception of experimental input of the vacancy formation energy. A previous investigation [49] reports that qEAM well reproduces experimental data such as thermal expansion and T_m up to about 3000K. We can now confirm that the qEAM melting curve is practically indistinguishable form the full *ab initio* result up to very high temperatures.

Additionally, we have shown that the CP can be calculated with reasonable effort directly from first principles. This can be accomplished as a general approach in materials where DFT is considered sufficiently accurate for the purpose of calculating thermodynamic properties. This work is a useful addition to the group of largely dispersed previous estimates, as it includes all relevant electronic contributions. This was achieved via carefully manipulated MD runs where a rough search in V-T phase space was performed for evidence of negative compressibility. Then several surrounding isotherms were calculated within the NVE ensemble while ensuring that the system remains in the single-component metastable fluid state during the data gathering stage around T_c. In Ta, the estimated density at the CP is nearly four times lower than in the solid at atmospheric pressure, yet the agreement with experiment is excellent, implying that inaccuracies of the DFT approach that typically occur at larger interatomic separation are not significant for CP determination. We have independently assessed the validity of various theoretical CP estimates, while ours is the closest to the only available experimental data, there is evidence that a two-parameter statistical model is usable to generate an approximate pressure field.

Furthermore, fitting the *ab initio* pressure field to virial power series results in an EOS that can be reliably extrapolated to ideal gas at temperatures at and above critical. Below T_c the available information is shown sufficient to estimate the P_{vap} curve. This is accomplished by combining the fluid and vapor pressure fields and determining the fluid-gas transition as the locus of points where these two pressures are equal. Here we can only determine the outset of the pure gas region as it moves rapidly into extreme simulation volumes due to strong attraction between Ta atoms. With our approach we cannot fully trace out this section of the binodal curve that separates two-component liquid-gas from pure gas phases below T_c since the simulation volumes required to explore a gas of highly attractive Ta atoms are unattainable by orders of magnitude. This leaves the pure gas phase as the only unresolved part of the V-T diagram. While the PR model turned out to be not accurate enough for this purpose, there is reason to believe another model could be of use; the density is so low in this region that much of it should be able to be well modelled even by an ideal gas.

Putting all these results together, a comprehensive multiphase EOS is established, deliverable as a table of Helmholtz free energies in over (V,T), as sketched in Fig. (4.13). The free energy field extends over unprecedented range of (V,T) space with volumes from 9 Å^3/atom (pressures over 7Mbars) to extremely dilute, and temperatures up to 20,000K. The free energy is accurately represented by piecewise analytical surfaces from which physical properties dependent on high order derivatives may be obtained.

In conclusion, we have constructed from first principles, an EOS of Ta that shows remarkable agreement with the available experimental data both in the very high density region (melting curve) and in the highly dilute region (critical point). In addition, it is well-behaved enough to provide properties that depend on high-order derivatives of the free energy. It is worth noting that while other approaches require

a specific strategy to obtain boundary features such as the melting curve, in our work the melting curve is a numerical by-product and its extraction can be largely automated. It is only part of a larger task of constructing a comprehensive EOS which makes its accuracy all the more striking. This EOS is accurate enough to serve as the basis for the analysis of previously published efforts in calculating sensitive thermodynamic features. Since our approach doesn't require a FF to be generated as long as the melting point is known, it may be the only available method for calculating EOS's in wide spans of phase space for many complex emerging materials whose dynamics is hard to parametrize.

Chapter 5

Retrospect and prospects

5.1 Summary

The computational work performed in this thesis represents the current state of the art in applying density functional theory to answer practical questions that arise in materials science. The many pitfalls involved in these tasks has been highlighted as well as some novel means to avoid them by increasing the reach of current simulation practices. We chiefly examined two problems where computational physics has great utility, developing an equation of state under conditions that make experimental measurement difficult and the effect of defects on the doping of a semiconductor, the exact mechanism being very demanding to determine in the lab.

Investigating the intrinsic doping of zinc phosphide involved both a careful consideration of the relevant thermodynamics as well as methods to broaden the reach of traditional DFT. Among these was a previously untried perturbative approach which allowed us to apply complex hybrid exchange-correlation functionals to systems larger than any that were calculated at the time. As a result of this careful study, we determined that the primary mechanism responsible for the lack of n-type doping were phosphorus interstitial defects. For zinc phosphide to be useful as a photovoltaic material these defects would have to be minimized; possible paths could include doping with electron donors which also inhibit phosphorus interstitials as well as crystal growth conditions which would increase the energy cost of interstitial defects (such

as zinc-rich environments or mechanically straining the crystal). In general, our work with zinc phosphide also produced an effective roadmap to study doping mechanisms in any material and should be particularly useful for low-symmetry crystals which often computationally expensive to simulate.

In developing an equation of state for tantalum we developed an efficient and accurate way to explore a large range of thermodynamic phase-space. This required extending the naturally zero-temperature framework of DFT to finite temperatures as well as determining where statistical mechanic modelling could complement it. By comparison with the available experimental and theoretical work, we are confident in having produced an equation of state with great precision and utility. Tantalum itself is a material of great interest for its corrosion resistance and use in high melting point alloys and this work should increase the accuracy of further studies involving multi-scale methods. Our theoretical work could also lead to improved force-fields for use in the molecular dynamic examination of both Ta and perhaps transition metals in general.

5.2 Future work

From here, there are many possible avenues that future work can take. For the dopant work in zinc phosphide there is a natural place to continue; now that we know the important defects to consider we can start looking for suitable dopants to obtain an n-type material. Modern methods are still primarily in the realm of developing a huge database of potential dopants and then data mining the result for the proper energetics and effect on bandstructure. On a personal level, it is discouraging that this scattershot approach is the most sophisticated means of solving this problem, but it also should not be surprising as this is what the computer really does best. The art in this work is to prioritize the most likely candidate materials as well as determine how to preserve the most relevant physics in our calculations (such as choice of which

functionals to use).

The 'perturbation extrapolation' method introduced with the zinc phosphide work is a very promising approach to increase the size of systems which can be calculated with more accurate and expensive functionals. There needs to be further work in formalizing the technique first in terms of assessing the validity of using a perturbative approach in a general system. This should also include a study of how to determine if the unit cell used to project the defect states onto constitutes a 'complete-enough' basis. There is also an open-question as to the requirements on the pseduopotentials used for this method to be generally useful.

Determining the doping mechanism of zinc phosphide was originally done with the purpose of finding a suitable photovoltaic material. This workflow can be applied generically to find other candidate materials as well as suitable dopants. However, defect studies are time-consuming to complete and it would very beneficial to screen candidate semiconductors and dopants before running these calculations. One means of accomplishing this is to determine the optical response of the crystal accurately and cull those materials which would be of no use as photovoltaics. While more advanced functionals do improve the reliability of DFT generated bandgaps and absorption properties, the GW method will continue to gain traction in this application as computing power increases. Determining the best way to utilize the GW technique in terms of starting conditions (since it is usually applied as a perturbative method) and which systems would benefit the most from its use are areas open to further study.

There is a common assumption that defect formation energies are largely independent of the effects of entropy (i.e. that all point defects in a system have roughly the same entropy benefit). Many common questions involve fine comparisons of the energies of, say a cation vacancy versus an anion interstitial defect, both of which have similar signatures experimentally. If there were a large vibrational entropy difference between these two types of defects the doping mechanism may be incorrectly

diagnosed. One could imagine that there may be structures where there are large entropy differences, especially in systems where certain types of defect wavefunctions are much more localized than others. We have started initial work investigating common prototypical crystal structures such as zincblend, rocksalt, etc. covering a wide range of electronegativity differences, ionic sizes, dielectric constants, etc. to try to characterize systems where defect entropy should be considered.

In summary, the promise of a virtual lab where one could arrange any combination of atoms and 'hit play' to obtain any physical characteristic precisely is still a point on the horizon in the field of computer simulation. Historically and even presently, simulation lends itself more readily to the explanation of physical phenomenon rather than prediction of truly new behavior - though this in itself can be very valuable insight into solving engineering problems with materials. The work in this thesis is right at the boundary of current research where, as with any field, there is as much art as science involved. However, as theories and methods become more refined many of the more novel calculations done for this thesis will become commonplace and evolve into more of a craft. Calculations which used to be exotic such as bandstructures are almost trivial in many cases now. Experiment will always be necessary, but there is a constantly growing region where simulation can complement lab work and researchers in either discipline are well-served to understand the capabilities of both.

Bibliography

[1] Schluter Hamann D. R. and Chiang C. *Phys. Rev. Let.*, 43:1494, 1979.

[2] Martin R. M. *Electronic Structure.* Cambridge University Press, 2004.

[3] Fumi. F. Bassani F. and Tosi M. P. *Highlights in Condensed Matter Theory.* 1985.

[4] Gilbert T. L. *Phys. Rev. B*, 12:2111, 1975.

[5] Janak J. F. *Phys. Rev. B*, 18:7165, 1978.

[6] Perdew J. P. Levy M. and Sahni V. *Phys. Rev. A*, 30:2745, 1984.

[7] John P. Perdew and Mel Levy. *Physical Review Letters*, 51:1884, 1983.

[8] Levy M. Perdew J. P., Parr R. G. and Balduz J. L. *Phys. Rev. Lett.*, 49:1691, 1982.

[9] Burke K. Perdew J. P. and Wang Y. *Phys. Rev. B*, 54:16533, 1996.

[10] Ernzerhof M. Perdew J. P. and Burke K. *Journal of Chem. Phys.*, 105:9982, 1996.

[11] Zunger A. Ihm J. and Cohen M. L. *Journal of Phys. Chem.*, 12:4409, 1979.

[12] L. Hedin. *Phys. Rev. A*, 139:796, 1965.

[13] Inkson J. C. *Many-body theory of solids: an introduction.* Plenum, 1983.

[14] Hedin L. and Lundqvist B. I. *Solid State Phys.*, 23:1, 1969.

[15] Holm B. *Phys. Rev. Let.*, 83:788, 1999.

[16] Gregory M. Kimball, Astrid M. Muller, Nathan S. Lewis, and Harry A. Atwater. *Applied Physics Letters*, 95:112103, 2009.

[17] Faa-Ching Wang, Alan L. Fahrenbruch, and Richard H. Bube. *Journal of Applied Physics*, 95:112103, 1982.

[18] M Bhushan and A Catalano. Polycrystalline zn3p2 schottky barrier solar cells. *Appl Phys Lett*, 38(1):39–41, 1981.

[19] W. Zdanowicz and Z. Henkie. *Bull. Acad. Polon. Sci. Ser. Sci. Chim.*, 12:729, 1964.

[20] Stephan Lany and Alex Zunger. *Physical Review Letters*, 98(4):045501, Jan 2007.

[21] Anderson Janotti and Chris G VandeWalle. Native point defects in zno. *Phys Rev B*, 76(16):165202, Jan 2007.

[22] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened coulomb potential. *The Journal of Chemical Physics*, 118(18):8207–8215, 2003.

[23] Stephan Lany and Alex Zunger. *Physical Review B*, 78:235104, 2008.

[24] Clas Persson, Yu-Jun Zhao, Stephan Lany, and Alex Zunger. *Physical Review B*, 72:035211, 2005.

[25] Stephan Lany and Alex Zunger. Accurate prediction of defect properties in density functional supercell calculations. *Model Simul Mater Sc*, 17(8):084002, Jan 2009.

[26] Stephan Lany. Semiconductor thermochemistry in density functional calculations. *Phys Rev B*, 78(24):245207, Dec 2008.

[27] G. Makov and M. C. Payne. *Physical Review B*, 51:4014, 1995.

[28] J IHM, A Zunger, and ML COHEN. Momentum-space formalism for the total energy of solids. *J Phys C Solid State*, 12(21):4409–4422, Jan 1979.

[29] Stephan Lany and Alex Zunger. Light- and bias-induced metastabilities in cu(in,ga)se-2 based solar cells caused by the (v-se-v-cu) vacancy complex. *J Appl Phys*, 100(11):113725, Jan 2006.

[30] T. S. Moss. *Proc. Phys. Soc. London*, 67:775, 1954.

[31] E. Burstein. *Phys. Rev.*, 93:632, 1954.

[32] A. Janotti, J. B. Varley, P. Rinke, N. Umezawa, G. Kresse, and C. G. VandeWalle. *Physical Review B*, 81:085212, 2010.

[33] Peter Agoston, Karsten Albe, Risto M. Nieminen, and Martti J. Puska. *Physical Review Letters*, 103:245501, 2009.

[34] J Heyd, JE Peralta, GE Scuseria, and RL Martin. Energy band gaps and lattice parameters evaluated with the heyd-scuseria-ernzerhof screened hybrid functional. *J Chem Phys*, 123(17):174101, Jan 2005.

[35] John P Perdew, Matthias Ernzerhof, and Kieron Burke. Rationale for mixing exact exchange with density functional approximations. *00219606*, 105(22):9982, 1996.

[36] A. Catalano and R.B. Hall. *J. Phys. Chem. Solids*, 41:635–640, 1979.

[37] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, Dec 1994.

[38] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758–1775, Jan 1999.

[39] G. Kresse and J. Furthmueller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15 – 50, 1996.

[40] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[41] von Stackelberg M and Paulus R. *Phys. Chem. B*, 28:427, 1935.

[42] L. G. Wang and Alex Zunger. Dilute nonisovalent (ii-vi)-(iii-v) semiconductor alloys: Monodoping, codoping, and cluster doping in znse-gaas. *Phys Rev B*, 68(12):125211, Sep 2003.

[43] Macdonald J. R. *Rev. Mod. Phys.*, 41:316, 1969.

[44] Greeff C. W. and Johnson J. D. *LA-13681-MS*, 2000.

[45] Boates B. and Bonev S. A. *Phys. Rev. Lett.*, 102:015701, 2009.

[46] Gillian M. J. Taioli S., Cazorla C. and Alfe D. *Phys. Rev. B*, 75:1214103, 2007.

[47] Habbal F. Li B. and Ortiz M. *Int. J. Numer. Math. Eng.*, 83:1541, 2010.

[48] Ravachandra G. Li B., Kidane A. and Ortiz M. *Int. J. Impact Eng.*, 42:25, 2012.

[49] Gulseren Mukherjee S. Cohen R. E. Strachan A., Cagin T. and Goddard W. A. III. *Model. Simul. Mater. Sci. Eng.*, 12:S445, 2004.

[50] Rudd R. E. S Moriarty J. A., Belak J. F.

[51] Allen M. P. and Tildesley D. J. *Computer Simulation of Liquids*, 1987.

[52] Ciccotti G. and Frenkel D. *Simulation of Liquids and Solids*, 1987.

[53] Frenkel D. and Smit B. *Understanding Molecular Simulations*, 2001.

[54] Marx D. and Hutter J. *Modern Methods and Alg. of Quant. Chem. Proc.*, 3:329, 2000.

[55] Kuhne T. *Journal of Chem. Phys.*, 2013.

[56] Marx D. and Hutter J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, 2009.

[57] Kolos W. *Adv. Quant. Chem.*, 5:99, 1970.

[58] Car R. and Parrinello M. *Phys. Rev. Lett.*, 55:2471, 1985.

[59] Smargiassi E. Pastore G. and Buda F. *Phys. Rev. A*, 44:6334, 1991.

[60] Blochl P. E. and Parrinello M. *Phys. Rev. B*, 45:9413, 1992.

[61] Frenkel D. and Smit B. *Understanding Molecular Simulation*, 1996.

[62] Sugino O. and Car R. *Phys. Rev. Lett.*, 74:1823, 1995.

[63] Kresse G. de Wijs G. and Gillan M. *Phys. Rev. B.*, 57:8223, 1998.

[64] Hong Q. and van de Walle A. *J. Chem. Phys.*, submitted.

[65] Peng D. Y. and Robinson D. B. *Ind. Eng. Chem., Fundam.*, 15:59, 1976.

[66] Kubo R.. *Journal of the Phys. Soc. of Japan*, 12:6, 1957.

[67] Mori-Sanchez P. Cohen A. J. and Yang W. *Science*, 321:792, 2008.

[68] Gillian M. J. Taioli S., Cazorla C. and Alfe D. *J. of Physics: Conf. Series*, 121:012010, 2008.

[69] Glosli J. N. Wu C. J., Soderlind P. and Klepeis J. E. *Nature Materials*, 8:223, 2009.

[70] van de Walle A. and Ceder G. *Journal of Phase Equilibria*, 23:348, 2002.

[71] Wang L. G. and van de Walle A. *Phys. Chem. Chem. Phys.*, 14:1529, 2012.

[72] Dinsdale A. T. *Calphad*, 15:317, 1991.

[73] Hultgren R. et. al. *Selected Values of the Thermodynamic Properties of the Elements*, 1973.

[74] Spurling T. H. Mason E. A. *The Virial Equation of State*, 1969.

[75] Bird R. B. Hirschfelder J. O., Curtiss C. F. *The Molecular Theory of Gases and Liquids*, 1964.

[76] Brown J. M. and Shaner J. W. *Shock Waves in Cond. Matter*, 1984.

[77] Gathers R. G. Shaner J. W. and Minichino C. *High Temp.-High Pressures*, 9:331, 1977.

[78] Terenovoi V. Ya *et al. High Temp.-High Press.*, 37:267, 2008.

[79] Iosilevski I. and Gryaznov V. *Proceedings of Int. Conf. HIF-2002*, 2010.

[80] Likalter A. A. *Physica A*, 311:137, 2002.

[81] Levashov P. R. *et al. AIP Conf. Proc.*, 505:89, 2000.

[82] Lomonosov I. V. *et al. AIP Conf. Proc.*, 620:111, 2002.

[83] Gathers G. R. *Rep. Prog. Phys.*, 49:341, 1986.

[84] Young D. A. and Alder B. J. *Phys. Rev. A*, 3:364, 1971.

[85] Fortov V.E. and Yacubov I.T. *Physics of Non-ideal Plasmas*, 1989.

[86] Moyano G. E. Wiebke J., Schwerdtfeger P. and Pahl E. *Chem. Phys. Lett.*, 514:164, 2011.

[87] Bohdansky J. and Schins H. E. J. *Journal of Phys. Chem.*, 71:215, 1967.

[88] Ishikawa T. Paradis P. and Yoda S. *Journal of Applied Phys.*, 97:053506, 2005.